THÈSE DE DOCTORAT

DE L'UNIVERSITÉ PIERRE ET MARIE CURIE

Ecole Doctorale 386

Sciences Mathématiques de Paris-Centre

Spécialité : Mathématiques

Option : Statistique

Présentée par Rawane SAMB

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE PIERRE ET MARIE CURIE

Sujet de la thèse :

CONTRIBUTION A L'ESTIMATION NONPARAMÉTRIQUE DE LA DENSITÉ DES ERREURS DE RÉGRESSION

Soutenue le 30 juin 2010 devant le jury composé de :

- M. Denis BOSQ — Examinateur
- M. Emmanuel GUERRE — Directeur de thèse
- Mme Ingrid VAN KEILEGOM — Rapporteur
- M. Christian FRANCQ — Rapporteur
- M. Benoît CADRE — Examinateur

Remerciements

Mes premiers remerciements s'adressent à mon directeur de thèse Emmanuel Guerre. Qu'il soit assuré de ma sincère gratitude pour la bienveillance et la grande disponibilité qu'il a toujours manifestées à mon égard durant toutes mes années de thèse. Il m'a beaucoup apporté par ses connaissances et sa rigueur scientifiques.

Je remercie très chaleureusement M. Denis Bosq, Professeur Emérite à l'université Pierre et Marie Curie, qui me fait l'honneur de présider le jury de cette thèse. Merci aussi à M. Christian Francq, Professeur à l'université Lille 3, et à Mme Ingrid Van Keilegom, Professeur à l'université catholique de Louvain en Belgique, d'avoir accepter de juger ce travail et d'en avoir été les rapporteurs. Leurs précieuses remarques et suggestions ont permis d'améliorer la qualité de cette thèse. Je suis également très reconnaissant envers M. Benoît Cadre, Professeur à l'ENS Cachan, qui m'a fait l'honneur d'accepter d'être membre du jury de cette thèse en tant qu'examinateur.

Je remercie aussi très vivement M. Paul Deheuvels, Professeur à l'université Pierre et Marie Curie, de m'avoir accueilli dans son DEA, puis dans son laboratoire. Je profite également de cette occasion pour remercier M. Philippe Saint-Pierre, Maître de Conférence du LSTA, pour son aide et ses encouragements, et tous les autres professeurs et maîtres de conférence du laboratoire.

Je suis extrêment reconnaissant envers M. Alain Chateauneuf et M. Jean Pierre Leca, respectivement Professeur et Maître de Conférence à l'université Paris 1 Panthéon-Sorbonne, pour leur sympathie et leurs qualités humaines. Ma reconnaissance va également à l'endroit de Mme Marie-Lou Margaria, Mme Brigitte Augarde et tout le personnel de l'université Paris 1.

Je souhaiterais également remercier les docteurs ou doctorants que j'ai connus à l'université Pierre et Marie Curie, notamment Mamadou Koné, Salim Bouzebda, Amadou Oury Diallo, Issam, Boris Labrador, François-Xavier, Hicham, Tarek, Mouhamed Cherfi, Choukri, Nabil, Layal, Lynda, Kaouthar, Lahcen Douge, Olivier Faugeras, Camille Sabbah, Olivier Bouaziz, Claire Coiffard, Clara, Emmanuel Onzon, Aurélie Fischer, Ousmane Bâ, entres autres, de m'avoir aidé ou montré leur sympathie au cours de ces années.

Je n'oublie pas mes amis Abass Sagna, Assane Diop, Massèye Gaye, Amadou Lamine Fall, Ange Toulougoussou, Serigne Touba Sall, Birame Diouf, Abdoulaye Sow, Mansor Sall, Gora Thiam et son épouse Fifi, Babacar Niang et Khourédia Ndiaye, Edwige Sophie Mendy, Lamine Bara Cissé et grand Bikèse à qui je fait part de ma reconnaissance et de ma gratitude. Merci également à Penda Sow, Cheikh Dia et son épouse Khady Kâne, Mamy Coumba Sanou Diouf, Astou Mbacké Fall et Mary Wade pour leur soutien et leur disponibilité.

Merci aussi à ma tante Khady Dia et à mes cousins Ndèye Niang, Ngomez, Ass, Khalifa et Assa. Je les remercie pour leur gentilesse, leur disponibilité et leur soutien qu'ils manifestent à mon égard depuis que je suis en France.

J'ai également une pensée très forte à mes parents Kiné Guèye, Mademba Samb, Bassirou Ndiaye et à toute ma famille. Je les remercie pour leur amour et leur soutien indéfectible et sans commune mesure à mon endroit. Je leur dédie ce travail. A ces remerciements et dédicaces, j'associe celle qui est devenue mon amie et ma douce moitié : je veux nommer mon épouse Rokhaya Dièye. Je lui suis reconnaissant pour son soutien moral.

Merci enfin à Massamba Wade, Rawane Wade, Mamadou Gassama, Samba Camara et à toutes celles et ceux qui m'ont aidé et accompagné pendant cette longue épreuve, et dont la place me manque ici pour les citer tous.

A ma mère Kiné Guèye,

A mon père Mademba Samb,

A mon oncle Bassirou Ndiaye,

A mes frères et soeurs,

A mon épouse Rokhaya Dièye.

Table des matières

Notations générales

Les notations suivantes seront utilisées dans les différents chapitres de cette thèse.

Ensembles, Nombres, Fonctions

Card (Ω) : Cardinal de l'ensemble Ω.

$\lfloor x \rfloor$: Partie entière du réel x.

$a \vee b$: Le maximum des réels a et b.

$a \wedge b$: Le minimum des réels a et b.

$\mathbb{1}_A$: Fonction indicatrice qui vaut 1 sur l'ensemble A et 0 ailleurs.

$f^{(k)}$: Dérivée k-ième de la fonction f.

Variables aléatoires

Soient X et Y deux variables aléatoires.

$\mathbb{E}(X)$: Espérance mathématique de X.

$\mathrm{Var}(X)$: Variance de X.

$\mathrm{Cov}(X, Y)$: Covariance de X et Y.

$\|X\|_p$: Norme L_p $(p \in]0, \infty[)$ de X définie par $\|X\|_p = \left(\mathbb{E}\left(|X|^p\right)\right)^{1/p}$, avec $\mathbb{E}\left(|X|^p\right) < \infty$.

Abréviations et Symboles

$:=$ Symbole utilisé pour la définition d'une quantité.

Soient $(a_n)_{n\geq 1}$ et $(b_n)_{n\geq 1}$ deux suites réelles.

$a_n = o(b_n)$, $n \to \infty$: Pour tout réel $\epsilon > 0$, on a $|a_n/b_n| \leq \epsilon$ pour n suffisamment grand.

$a_n = O(b_n)$, $n \to \infty$: Il existe un réel $C > 0$ tel que $|a_n/b_n| \leq C$ pour n suffisamment grand.

$a_n \asymp b_n$, $n \to \infty$: $a_n = O(b_n)$ and $b_n = O(a_n)$ pour n suffisamment grand.

Chapitre 1

Introduction Générale

1.1 Présentation du sujet

Soit $(X_1, Y_1), \ldots, (X_n, Y_n)$ un échantillon de variables aléatoires indépendentes et identiquement distribuées (i.i.d), de même loi que (X, Y). On suppose que Y est une variable univariée à valeurs dans $\mathbb{R}$, et que X désigne une variable explicative multivariée prenant ses valeurs dans $\mathbb{R}^d$, $d \geq 1$. Soit $m(\cdot)$ l'espérance conditionnelle de Y sachant X, de telle sorte que le modèle de régression relatif à X et Y s'écrit

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1.1}$$

où les erreurs ε_i sont supposées être des variables aléatoires i.i.d, indépendantes des X_i, de même loi que ε satisfaisant en particulier $\mathbb{E}[\varepsilon] = 0$.

Dans ce mémoire de thèse, nous étudions l'estimation nonparamétrique de la densité f de l'erreur du modèle (1.1). Cette estimation de la densité de l'erreur de régression est un important outil descriptif permettant de comprendre le comportement des résidus, et de faire des tests d'hypothèses sur la distribution des erreurs du modèle ou sur la fonction de régression. On pourra consulter, par exemple, Ahmad et Li (1997), Dette et *al.* (2002), Neumeyer et *al.* (2005), pour le test de symétrie de la distribution des erreurs de régression ; Akritas et Van Keilegom (2001), Cheng et Sun (2008), pour des tests d'ajustement sur la loi des résidus ; Gozalo et Linton (2001), Dette et von Lieres und Wilkau (2001), Neumeyer et Van Keilegom (2010), pour le test sur l'additivité de la fonction de régression. Notons aussi que l'estimation de f peut trouver son importance dans la prévision de Y_{n+1} à partir de X_{n+1}. En effet, on peut prédire Y_{n+1} par l'estimateur du mode conditionnel $\text{mod}\,(x)$ de Y_{n+1} sachant que $X_{n+1} = x$, puisque $\text{mod}\,(x) = m\,(x) + \arg\max_{\epsilon \in \mathbb{R}} f(\epsilon)$. Le fait d'estimer f est également très important dans la détermination d'un intervalle de prédiction pour Y_{n+1}, ce qui nécessite d'estimer des quantiles de la loi f. L'estimation de f peut aussi servir à estimer la loi de la variable Y, comme relaté dans Escanciano et Jacho-Chavez (2010). Enfin cette estimation de la loi des résidus peut être utile pour la construction d'estimateurs nonparamétriques de la densité et de la fonction de hazard de Y sachant X. Voir Van Keilegom et Veraverbeke (2002).

Pour estimer la densité f des résidus du modèle (1.1), une première approche consiste à noter que la densité f se déduit de la densité $\varphi\,(\cdot|x)$ de Y sachant que $X = x$. Plus précisémment, on a la relation

$$f(\epsilon) = \varphi\,(\epsilon + m(x)|x)\,. \tag{1.2}$$

Suivant cette idée, on peut donc en théorie déduire un estimateur de $f(\epsilon)$ à partir d'une estimation de $\varphi\,(y|x)$ et de $m(x)$. Cette approche est cependant sujette au "fléau de la dimension" : l'estimation de $\varphi\,(y|x)$ ne peut se faire qu'avec une vitesse très lente lorsque la dimension de x est élevée. Les approches proposées dans cette thèse visent à "déconditionner"dans l'expression (1.2) de $f(\epsilon)$. En effet, la relation (1.2) entraîne que

$$f(\epsilon) = \int \varphi\,(\epsilon + m(x)|x)\,g(x)dx, \tag{1.3}$$

où $g(x)$ désigne la densité de X. Cette nouvelle formule suggère que le "fléau de la dimension"n'est peut être pas aussi important que le laissait penser la première approche basée sur les estimations de $f\,(y|x)$ et de $m(x)$. Deux stratégies sont mises en oeuvre dans cette thèse pour essayer d'éviter le "fléau de la dimension". La première consiste à estimer nonparamétriquement chaque résidu ε_i par $\widehat{\varepsilon}_i = Y_i - \widehat{m}_n(X_i)$, où $\widehat{m}_n(\cdot)$ désigne un estimateur nonparamétrique de la fonction de régression $m(\cdot)$. La seconde consiste à procéder comme dans (1.3), et à étudier l'estimateur

$$\widehat{f}_n(\epsilon) = \int \widehat{\varphi}_n\,(\epsilon + \widehat{m}_n(x)|x)\,\widehat{g}_n(x)dx,$$

où $\widehat{\varphi}_n(\cdot|x)$ et $\widehat{g}_n(x)$ désignent respectivement des estimateurs nonparamétriques de $\varphi(\cdot|x)$ et $g(x)$.

Le problème de l'estimation de la densité des résidus d'un modèle régression est un cas particulier d'un problème plus général : l'estimation d'un paramètre d'intérêt en présence d'un paramètre de nuisance. Dans notre cadre, qui se focalise sur l'estimation de la distribution des résidus, la densité des résidus $f(\cdot)$ est le paramètre d'intérêt, et la fonction de régression $m(\cdot)$ le paramètre de nuisance. La présence de ce paramètre de nuisance dans le modèle va influencer l'estimation du paramètre d'intérêt. Dans le cas paramétrique, considérons, par exemple, un échantillon $Z, Z_1, \ldots, Z_n$ de variables aléatoires indépendantes et identiquement distribuées, de densité $f\,(z|\theta,\eta)$, où θ est le paramètre d'intérêt et η le paramètre de nuisance. Une quantité centrale liée à ces deux paramètres est la matrice d'information de Fischer

$$I(\eta,\theta) = \mathrm{Var}\left[\nabla f(z|\eta,\theta)\right],$$

où $\nabla f(z|\eta,\theta)$ est le gradient de $f(z|\eta,\theta)$ par rapport à η et θ défini par

$$\nabla f(z|\eta,\theta) = \begin{bmatrix} \frac{\partial}{\partial\eta} f(z|\eta,\theta) \\ \frac{\partial}{\partial\theta} f(z|\eta,\theta) \end{bmatrix}.$$

La matrice $I(\eta,\theta)$ s'écrit sous la forme d'une matrice en blocs

$$I(\eta,\theta) = \begin{bmatrix} I_{\eta\eta} & I_{\eta\theta} \\ I_{\theta\eta} & I_{\theta\theta} \end{bmatrix},$$

où

$$I_{\theta\theta} = \mathrm{Var}\left[\frac{\partial}{\partial\theta} f(z|\eta,\theta)\right], \quad I_{\eta\eta} = \mathrm{Var}\left[\frac{\partial}{\partial\eta} f(z|\eta,\theta)\right].$$

L'inégalité de Fréchet-Darmois-Cramer-Rao (Borovkov 1987, page 156) montre que l'inverse de la matrice d'information de Fischer, $I^{-1}\,(\eta,\theta)$, est, au sens de l'ordre sur les matrices, la plus petite matrice de variance possible pour les estimateurs sans biais de (η,θ). Cette borne $I^{-1}\,(\eta,\theta)$ est atteinte par les estimateurs du maximum de vraisemblance, comme le rappelle le théorème suivant.

Theorème 1.1. (Borovkov 1987, page 229)
Soit $(\widehat{\eta}_n, \widehat{\theta}_n)$ *un estimateur du maximum de vraisemblance de* (η, θ)*. Sous certaines conditions de régularité, on a la convergence asymptotique suivante :*

$$\sqrt{n}\begin{pmatrix}\widehat{\eta}_n - \eta \\ \widehat{\theta}_n - \theta\end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, I^{-1}(\eta, \theta)\right).$$

La formule du calcul de l'inverse d'une matrice en blocs appliquée à $I(\eta, \theta)$ permet de voir que

$$I^{-1}(\eta, \theta) = \begin{bmatrix} I^{\eta\eta} & I^{\eta\theta} \\ I^{\theta\eta} & I^{\theta\theta}\end{bmatrix},$$

avec

$$I^{\theta\theta} = \left(I_{\theta\theta} - I_{\theta\eta} I_{\eta\eta}^{-1} I_{\eta\theta}\right)^{-1}.$$

Du théorème précédent, on déduit la loi limite de l'estimateur du paramètre d'intérêt θ.

Corollaire 1.1. *Sous les conditions du théorème précédent, on a la convergence asymptotique*

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, I^{\theta\theta}\right).$$

La matrice $I^{\theta\theta}$ s'interprète, grâce à l'inégalité de Fréchet-Darmois-Cramer-Rao, comme étant la meilleure variance possible pour un estimateur sans biais de θ, η étant inconnu. Puisque $I_{\theta\eta} I_{\eta\eta}^{-1} I_{\eta\theta}$ est semi-positive, la formule de $I^{\theta\theta}$ suggère que $I^{\theta\theta}$ est, au sens de l'ordre sur les matrices symétriques, plus grande que $I_{\theta\theta}^{-1}$ sauf si $I_{\eta\theta} = 0$, condition indiquant que les estimateurs du maximum de vraisemblance de θ et η sont asymptotiquement indépendants. Comme la variance asymptotique de l'estimateur de θ quand η est connu est $I_{\theta\theta}^{-1}$, cette différence entre $I^{\theta\theta}$ et $I_{\theta\theta}^{-1}$ mesure la perte (en terme d'efficacité) du fait que η soit inconnu quand on veut estimer θ.

Une autre situation proche du problème de l'estimation de la densité des résidus est l'estimation de la fonction de répartition lorsque des paramètres sont inconnus. Considérons, par exemple, un échantillon $X_1, \ldots, X_n$ de variables aléatoires i.i.d de fonction de répartition commune $F(x, \theta)$, où $\theta \in \mathbb{R}$. Pour un estimateur $\widehat{\theta}_n$ de θ, on définit la fonction empirique associée

$$\widehat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left(F(X_i, \widehat{\theta}_n) \leq t\right), \quad t \in [0, 1].$$

Cette fonction de répartition empirique joue un rôle important pour les tests d'adéquation du modèle considéré. En effet, $\widehat{F}_n(t)$ doit être proche de t si le modèle est correctement choisi. Considérons, par exemple, le modèle de translation

$$X_i = \theta + \varepsilon_i, \quad i = 1, \ldots, n,$$

où les résidus ε_i sont de distribution commune ψ. On a $F(x, \theta) = \psi(x - \theta)$. Pour ce modèle paramétrique, on a

$$F(X_i, \widehat{\theta}_n) = \psi(X_i - \widehat{\theta}_n) = \psi(\widehat{\varepsilon}_i),$$

où $\widehat{\varepsilon}_i$ est le résidu estimé $X_i - \widehat{\theta}_n$. En conséquence, on a

$$\widehat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left(\psi(\widehat{\varepsilon}_i) \leq t\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left(\widehat{\varepsilon}_i \leq \psi^{-1}(t)\right).$$

La relation ci-dessus montre donc que $\widehat{F}_n(t)$ est, à une transformation de t près, la fonction de répartition empirique des résidus $\widehat{\varepsilon}_i$. Le processus empirique associé à $\widehat{F}_n$ est

$$\widehat{y}_n(t) = n^{1/2}\{\widehat{F}_n(t) - t\}, \quad t \in [0, 1].$$

Ce processus a été étudié par Durbin (1973) qui obtint le résultat suivant.

Theorème 1.2. *Soit $\widehat{\theta}_n$ un estimateur de θ tel que*

$$n^{1/2}(\widehat{\theta}_n - \theta) = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \ell(x_i, \widehat{\theta}_n) + o_{\mathbb{P}}(1),$$

où ℓ est une fonction mesurable telle que $\mathbb{E}\left[\ell(X_1, \theta)\right] = 0$. Pour tout $t \in [0, 1]$, on définit la fonction $g(t)$ par

$$g(t) = g(t, \theta) = \frac{\partial F(x, \theta)}{\partial \theta} \,|_{x=Q(t,\theta)}, \quad Q(t, \theta) = \inf\{z : F(z, \theta) = t\},$$

et on pose

$$\begin{aligned} h(t) &= h(t, \theta) = \int_{-\infty}^{Q(t,\theta)} \ell(x, \theta) dF(x, \theta), \\ L(\theta) &= \mathbb{E}\left[\ell^2(X_1, \theta)\right]. \end{aligned}$$

Alors sous des conditions de régularité, le processus $\{\widehat{y}_n(t), 0 \leq t \leq 1\}$ converge asymptotiquement en distribution vers un processus gaussien $\{y(t), 0 \leq t \leq 1\}$, de moyenne nulle et de fonction de covariance

$$\mathrm{Cov}\left(y(t_1), y(t_2)\right) = \min(t_1, t_2) - t_1 t_2 - h(t_1)g(t_2) - h(t_2)g(t_1) + g(t_1)L(\theta)g(t_2),$$

On note que cette fonction de covariance dépend de la fonction de répartition $F(\cdot, \theta)$ inconnue. Donc la distribution asymptotique obtenue pour le processus $\widehat{y}_n(t)$ est différente de la loi limite obtenue pour le processus empirique usuel (qui suppose θ connu),

$$y_n(t) = n^{1/2}\{F_n(t) - t\}, \quad F_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left(F(X_i, \theta) \leq t\right).$$

En effet, il a été démontré que le processus $\{y_n(t), 0 \leq t \leq 1\}$ converge asymptotiquement vers un pont Brownien. Voir, par exemple, le livre de Billinsgley (1968, p.109).

La suite de cette introduction générale donne des exemples d'estimation de paramètres dans le cas d'un modèle de régression $Y = m(X) + \sigma(X)\varepsilon$. Ces exemples seront donnés selon que le paramètre de nuisance, ici la fonction de régression $m(\cdot)$, est paramétrique ou non.

1.2 Estimation de la fonction de répartition des résidus d'un modèle linéaire

On considère le modèle linéaire

$$Y_i = \theta^\top X_i + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1.4}$$

où les erreurs ε_i sont i.i.d de fonction de répartition commune F. Les variables X_i sont supposées non aléatoires. Soit $\widehat{\theta}_n$ un M-estimateur de θ (Consulter, par exemple, Huber 1964, 1981). On

s'intéresse au comportement asymptotique de la fonction de répartition empirique $\widehat{F}_n$ des résidus estimés $\widehat{\varepsilon}_i = Y_i - X_i^\top \widehat{\theta}_n$,

$$\widehat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1} \left(\widehat{\varepsilon}_i \le t \right), \quad t \in \mathbb{R},$$

lorsque la dimension p des régresseurs peut dépendre de la taille n de l'échantillon. Ce problème a été étudié par Portnoy (1986) et Mammen (1996). Portnoy (1986) obtient le développement

$$n^{1/2} \left(\widehat{F}_n(t) - F_n(t) \right) = \frac{f(t)}{n^{1/2}} \sum_{i=1}^{n} X_i^\top \left(\widehat{\theta}_n - \theta \right) + o_{\mathbb{P}}(1), \tag{1.5}$$

où $F_n(t)$ est la fonction de répartition empirique basée sur les vrais résidus. Puis il montre que ce développement (1.5) n'a lieu que si $p^2/n = O(1)$ lorsque n tend vers l'infini. Mammen (1996) s'intéresse au comportement asymptotique de $\widehat{F}_n$ lorsque p^2/n est divergente. Il considère un M-estimateur $\widehat{\theta}_\psi$ tel que

$$\widehat{\theta}_\psi - \theta - \sum_{i=1}^{n} X_i G(\varepsilon_i) = O_{\mathbb{P}} \left(\frac{p^2}{n} \right)^{1/2}, \quad G(t) = \frac{\psi(t)}{\mathbb{E}\psi^{(1)}(\varepsilon_i)}, \quad t \in \mathbb{R}, \quad \mathbb{E}\left[G(\varepsilon_i) \right] = 0,$$

où ψ est une fonction dérivable et croissante. Sous des conditions de régularité, Mammen montre que pour tout $0 < C < \infty$,

$$\sup_{|t| \le C} \left| n^{1/2} \left(\widehat{F}_n(t) - F_n(t) \right) - \Delta_n(t) \right| = o_{\mathbb{P}}(1), \tag{1.6}$$

où, si f désigne la densité des résidus,

$$\Delta_n(t) = \frac{f(t)}{n^{1/2}} \sum_{i=1}^{n} \left[X_i^\top \left(\widehat{\theta}_n - \theta \right) \right] + \frac{f(t)p}{n^{1/2}} \left[G(t) + \frac{f^{(1)}(t)}{2f(t)} \mathbb{E} G^2(\varepsilon_1) \right].$$

Dans le résultat (1.5) de Portnoy, il n'y a pas d'influence asymptotique de l'estimation des résidus sur l'estimateur de la distribution $F(t)$ lorsque

$$\frac{1}{n^{1/2}} \sum_{i=1}^{n} X_i^\top (\widehat{\theta}_n - \theta) = \frac{1}{n} \sum_{i=1}^{n} X_i^\top \sqrt{n}(\widehat{\theta}_n - \theta) = o_{\mathbb{P}}(1).$$

Donc, puisque $\sqrt{n}(\widehat{\theta}_n - \theta) = O_{\mathbb{P}}(1)$, sous des hypothèses de régularité usuelles, la condition ci-dessus est réalisée lorsque $\mathbb{E}[X] = 0$, d'après la loi des grands nombres. Pour le résultat (1.6) de Mammen, il y a un effet de l'estimation des résidus. En effet, le terme $\Delta_n(t)$ ne peut pas être négligeable puisque p^2/n diverge.

L'estimation de la distribution des résidus a aussi été étudiée dans le cadre des modèles autoregressifs linéaires. Dans le autorégressif d'ordre 1 AR(1), on observe les variables aléatoires $X_0, X_1, \ldots, X_n$ telles que

$$X_i = \rho X_{i-1} + \varepsilon_i, \quad 1 \le i \le n,$$

où ρ désigne un paramètre réel, et les ε_i des variables aléatoires indépendantes et identiquement distribuées (i.i.d) de densité de probabilité f définie sur $\mathbb{R}$. Pour estimer la fonction de répartition F des résidus, on estime d'abord les résidus ε_i par $\widehat{\varepsilon}_i = X_i - \widehat{\rho}_n X_{i-1}$, $\widehat{\rho}_n$ pouvant être obtenu par la méthode des moindres carrées ordinaires. Le théorème suivant obtenu par Koul (1992) donne une idée sur l'effet de l'estimation des résidus sur la loi limite de l'estimateur de F.

Theorème 1.3. *Soit $\widehat{\rho}_n$ un estimateur de ρ tel que $n^{1/2}(\widehat{\rho}_n - \rho) = O_{\mathbb{P}}(1)$. Alors sous une hypothèse d'ergodicité de la famille $\{\varepsilon_i, 1 \leq i \leq n\}$, et sous d'autres hypothèses convenables, on a*

$$\sup_{x\in\mathbb{R}} \left| n^{1/2} \left[F_n\left(x, \widehat{\rho}_n\right) - F_n\left(x, \rho\right) \right] \right| = o_{\mathbb{P}}(1).$$

Le résultat de ce théorème montre que l'estimation du paramètre ρ n'a pas un effet asymptotique sur l'estimation de la fonction de répartition F des résidus du modèle précédent. Ceci vient de ce que le modèle AR(1) est très proche du modèle linéaire (1.4), les variables X_i étant de moyenne nulle.

1.3 Estimation des moments d'une fonctionnelle de l'erreur

La fonction de répartition correspond à un moment particulier, le moment de la fonction $\mathbb{1}(\varepsilon \leq t)$. Müller, Schick et Wefelmeyer (2004) ont étudié le cas plus général d'un moment $\mathbb{E}h(\varepsilon)$, mais en supposant que h est différentiable. Leur cadre d'étude est le modèle de régression non-paramétrique $Y = m(X) + \varepsilon$, où ε est indépendante de X. La fonction h est supposée connue. Le modèle est basé sur un échantillon d'observations i.i.d $(X_1, Y_1), \ldots, (X_n, Y_n)$ de même loi que (X, Y). Les résidus ε_i sont estimés par $\widehat{\varepsilon}_i = Y_i - \widehat{m}(X_i)$, où $\widehat{m}$ est un estimateur non paramétrique de m. Les auteurs proposent d'estimer $\mathbb{E}[h(\varepsilon)]$ par $\widehat{H}_n = n^{-1}\sum_{i=1}^n h(\widehat{\varepsilon}_i)$. Sous des conditions de régularité, ces auteurs montrent que $\widehat{H}_n$ est un estimateur efficace de $\mathbb{E}[h(\varepsilon)]$ tel que

$$\widehat{H}_n = \frac{1}{n}\sum_{i=1}^n \left[h(\varepsilon_i) - \mathbb{E}[h^{(1)}(\varepsilon)]\varepsilon_i \right] + o_{\mathbb{P}}(n^{-1/2}).$$

En conséquence, la quantité $n^{1/2}[\widehat{H}_n - \mathbb{E}h(\varepsilon)]$ converge asymptotiquement vers une distribution normale de moyenne nulle et de variance

$$\tau_*^2 = \mathbb{E}\left[\left(h(\varepsilon) - \mathbb{E}h(\varepsilon) - \mathbb{E}[h^{(1)}(\varepsilon)]\varepsilon \right)^2 \right].$$

Un aspect surprenant de ce résultat est que, pour certaines fonctions h, la variance asymptotique τ_*^2 de $\widehat{H}_n$ est plus petite que la variance asymptotique τ^2 de l'estimateur $H_n = n^{-1}\sum_{i=1}^n h(\varepsilon_i)$ basé sur les vrais résidus. En effet, supposons, par exemple, que les résidus suivent une loi normale de moyenne nulle et variance égale à σ^2. Pour simplifier, on suppose que $\sigma^2 = 1$. Puisque la variance asymptotique de l'estimateur H_n est égale $\tau^2 = \mathbb{E}[(h(\varepsilon) - \mathbb{E}h(\varepsilon))^2]$, on a $\tau_*^2 < \tau^2$ si et seulement si

$$0 < \mathbb{E}[h^{(1)}(\varepsilon)] < 2\mathbb{E}[\varepsilon h(\varepsilon)] \;\text{ ou }\; 2\mathbb{E}[\varepsilon h(\varepsilon)] < \mathbb{E}[h^{(1)}(\varepsilon)] < 0. \tag{1.7}$$

De plus, dans le cas où la variable ε suit une loi normale de variance $\sigma^2 = 1$, on a, sous des hypothèses convenables, $\mathbb{E}[h^{(1)}(\varepsilon)] = \mathbb{E}[\varepsilon h(\varepsilon)]$. En conséquence, la première double inéqualité dans (1.7) est vérifiée si $\mathbb{E}[h(\varepsilon)\varepsilon] < 0$, alors que la seconde double inéqualité dans (1.7) est satisfaite lorsque $\mathbb{E}[h(\varepsilon)\varepsilon] > 0$. Cette dernière condition est par exemple vérifiée lorsque $h(z) = z^3$, avec ε suivant une loi normale centrée réduite. Ce qui, dans un tel cas, entraîne que $\tau_*^2 < \tau^2$. Un tel paradoxe s'explique par le fait que l'estimateur $\widehat{H}_n$ utilise mieux le fait que les résidus ε_i sont de moyenne nulle.

1.4 Estimation nonparamétrique de la densité de l'erreur dans un modèle autorégressif non linéaire

Fu et Yang (2008) étudient la distribution asymptotique d'un estimateur à noyau de la densité de l'erreur dans un modèle AR(p) non linéaire. Ce modèle est de la forme

$$X_i = g_\theta(X_{i-1}, \dots, X_{i-p}) + \varepsilon_i, \quad i \geq 1,$$

où $\{X_i, i \in \mathbb{Z}\}$ est strictement stationnaire, et $\theta = (\theta_1, \dots, \theta_q)^\top \in \mathbb{R}^q$. Les ε_i sont i.i.d, de densité f, avec une moyenne nulle et une variance σ^2. On suppose également que les résidus ε_i sont indépendantes de la famille $(X_{i-1}, \dots, X_{i-p})$. Pour un estimateur $\widehat{\theta} = (\widehat{\theta}_1, \dots, \widehat{\theta}_q)^\top$, on estime les résidus ε_i par

$$\widehat{\varepsilon}_i = X_i - g_{\widehat{\theta}}(X_{i-1}, \dots, X_{i-p}), \quad i \geq 1.$$

En utilisant ces résidus empiriques, Fu et Yang estiment nonparamétriquement la densité f par

$$\widehat{f}_n(t) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{\widehat{\varepsilon}_i - t}{h_n}\right), \quad t \in \mathbb{R},$$

où (h_n) est une suite de réels positifs tendant vers zero quand n tend vers l'infini, et K une fonction noyau définie sur $\mathbb{R}$. En désignant par

$$f_n(t) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{\varepsilon_i - t}{h_n}\right), \quad t \in \mathbb{R},$$

l'estimateur nonparamétrique de f basé sur les vrais résidus, Fu et Yang obtiennent le résultat suivant.

Theorème 1.4. Fu et Yang (2008)
Supposons qu'il existe un réel $C_1 > 0$ tel que l'estimateur $\widehat{\theta}$ vérifie, avec une probabilité égale à 1,

$$\lim_{n\to\infty} \sup \sqrt{\frac{n}{\log\log n}} \|\widehat{\theta} - \theta\| \leq C_1, \tag{1.8}$$

où $\|x\|^2 = \sum_{j=1}^{q} x_j^2$ pour tout $x = (x_1, \dots, x_q)^\top \in \mathbb{R}^q$. On suppose également que la fenêtre h_n satisfait

$$h_n \to 0, \quad \lim_{n\to\infty} \frac{n^{1/2} h_n^{5/2}}{\log\log n} = \infty. \tag{1.9}$$

Alors sous certaines conditions de régularité, on a la convergence en distribution suivante :

$$\frac{1}{\sqrt{\mathrm{Var} f_n(t)}} \left(\widehat{f}_n(t) - \mathbb{E} f_n(t)\right) \xrightarrow{d} \mathcal{N}(0,1),$$

où $\mathcal{N}(0,1)$ désigne la loi normale centrée réduite.

La condition (1.8) est satisfaite par un estimateur du maximum de vraisemblance sous certaines conditions proposées par Klimko et Nelson (1978).

Il a été démontré dans la littérature statistique que $n^{-1/5}$ est l'ordre de la fenêtre optimale pour l'estimation nonparamétrique de la densité d'une variable aléatoire réelle ζ à partir d'un échantillon de variables aléatoires i.i.d $\zeta_1, \zeta_2, \dots, \zeta_n$. Pour ce résultat, on peut, par exemple, se référer aux ouvrages de Bosq et Lecoutre (1987), Scott (1992), Wand et Jones (1995). On note que dans le cadre du théorème précédent, la condition (1.9) ne peut pas vérifiée lorsque h_n est d'ordre $n^{-1/5}$, mais que tous les ordres $n^{-(1/5)+\epsilon}$, $\epsilon > 0$, qui s'en approchent sont possibles.

1.5 Estimation de la loi des résidus en régression nonparamétrique

L'étude de l'estimation nonparamétrique d'une distribution de l'erreur dans un modèle de régression nonparamétrique occupe une place importante dans la litterature statistique. En effet, plusieurs résultats inhérents à ce type d'estimation ont été obtenus au début de cette décennie. On peut citer, par exemple, Akritas et Van Keilegom (2001) dans le cadre de l'estimation non paramétrique de la fonction de répartition de l'erreur d'un modèle de régression hétéroscédastique, puis Efromovich (2005, 2007) et Cheng (2005) pour l'estimation nonparamétrique de la densité des résidus d'un modèle de régression homoscédastique. Plus récemment, Wang, Brown, Cai et Levine (2008) se sont intéressés à l'étude de l'influence de la fonction moyenne conditionnelle, supposée inconnue, sur l'estimation de la variance conditionnelle des résidus dans le cas d'un modèle de régression hétéroscédastique.

1.5.1 Estimation de la fonction de répartition des résidus dans un modèle de régression hétéroscédastique

Akritas et Van Keilegom (2001) proposent un estimateur nonparamétrique de la fonction de répartition F de l'erreur ε dans le modèle de régression hétéroscedastique $Y = m(X) + \sigma(X)\varepsilon$, où ε est indépendante de X, et m et σ des fonctions "lisses"satisfaisant quelques conditions de régularité. L'estimateur $\widehat{F}_n$ de F_ε est basé sur l'estimation nonparamétrique des résidus $\varepsilon_i = (Y_i - m(X_i))/\sigma(X_i)$, où $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$ désignent un échantillon d'observations indépendantes et identiquement distribuées. Pour l'estimation de ces résidus, Akritas et Van Keilegom écrivent $m(x)$ sous la forme

$$m(x) = \int_0^1 F^{-1}(s|x)ds, \tag{1.10}$$

où $F^{-1}(s|x) = \inf\{y \in \mathbb{R} : F(y|x) \geq s\}$, $F(y|x) = \mathbb{P}(Y \leq y|x)$. On note que si la fonction F est continue, le changement de variable $s = F(u|x)$ dans (1.10) entraîne

$$\int_0^1 F^{-1}(s|x)ds = \int_{\mathbb{R}} udF(u|x) = \mathbb{E}\left[Y|X = x\right] = m(x).$$

Pour l'estimation de F_ε, les auteurs estiment dans un premier temps $F(y|x)$ par l'estimateur de Stone (1977)

$$\widetilde{F}(y|x) = \sum_{i=1}^n W_i(x, a_n)\mathbb{1}(Y_i \leq y),$$

où les $W_i(x, a_n)$ sont les poids de Nadaraya-Watson (1964) définis par

$$W_i(x, a_n) = \frac{K\left(\frac{X_i - x}{a_n}\right)}{\sum_{j=1}^n K\left(\frac{X_j - x}{a_n}\right)},$$

avec K désignant une fonction noyau, et a_n une fenêtre tendant vers 0 lorsque n tend vers l'infini. Dans un deuxième temps, Akritas et Van Keilegom estiment $m(x)$ et $\sigma^2(x)$ par

$$\widehat{m}(x) = \int_0^1 \widetilde{F}^{-1}(s|x)ds, \quad \widehat{\sigma}^2(x) = \int_0^1 \widetilde{F}^{-1}(s|x)^2 ds - \widehat{m}^2(x).$$

Il convient de signaler à nouveau que le changement de variable $s = \widetilde{F}_n^{-1}(y|x)$ entraîne

$$\int_0^1 \widetilde{F}^{-1}(s|x)ds = \sum_{i=1}^n Y_i W_i(x, a_n),$$

ce qui correspond à l'estimateur de Nadaraya-Watson (1964) classique.

Avec l'aide de ces estimateurs de $m(x)$ et $\sigma(x)$, on estime chaque résidu ε_i par $\widehat{\varepsilon}_i = (Y_i - \widehat{m}(X_i))/\widehat{\sigma}(X_i)$. L'estimateur de $F_\varepsilon(t)$ basé sur les résidus estimés est alors défini par $\widehat{F}_\varepsilon(t) = n^{-1}\sum_{i=1}^n \mathbb{1}(\widehat{\varepsilon}_i \le t)$. Pour la détermination de la loi limite de cet estimateur, Akritas et Van Keilegom proposent d'abord un développement asymptotique de $\widehat{F}_\varepsilon(t)$. Ce développement est donné par le théorème suivant.

Theorème 1.5. *On suppose que la fonction de répartition F_X de X est trois fois dérivable sur le support $\mathcal{X}$ de X, et que et la densité f_X de X vérifie $\inf_{x\in\mathcal{X}} f_X(x) > 0$. On suppose également que les fonctions $m(\cdot)$ et $\sigma(\cdot)$ sont deux fois continûment dérivables sur $\mathcal{X}$ et que $\inf_{x\in\mathcal{X}} \sigma(x) > 0$. Alors pour tout $t \in \mathbb{R}$, on a*

$$\widehat{F}_\varepsilon(t) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}\left(\frac{Y_i - m(X_i)}{\sigma(X_i)} \le t\right) - F_\varepsilon(t) + \frac{1}{n}\sum_{i=1}^n \varphi(X_i, Y_i, t) + \beta_n(t) + o_{\mathbb{P}}(n^{-1/2}) + o_{\mathbb{P}}(a_n^2),$$

où

$$\begin{aligned} \varphi(x,y,t) &= -\frac{f_\varepsilon(t)}{\sigma(x)}\int \left[\mathbb{1}(y \le v) - F(v|x)\right]\left[1 + t\frac{v - m(x)}{\sigma(x)}\right] dv, \\ \beta_n(t) &= \frac{a_n^2 \mu_K}{2}\int \frac{\partial^2}{\partial x^2}\mathbb{E}\left[\varphi(x, Y, t)|u\right]|_{x=u}\, dF_X(u), \end{aligned}$$

avec f_ε désignant la densité de ε, μ_K une constante qui dépend de K, et F_X la fonction de répartition de X.

De ce théorème, Akritas et Van Keilegom déduisent le corollaire suivant qui donne un résultat de convergence asymptotique du processus $n^{1/2}(\widehat{F}_\varepsilon(t) - F_\varepsilon(t))$. Ce résultat étend les travaux de Durbin (1973) et Loynes (1980) concernant la loi asymptotique d'un estimateur de la fonction de répartition des résidus basé sur des paramètres estimés.

Corollaire 1.2. *Supposons que le Théorème 1.5 est vérifié.*
(i) *Si $na_n^4 \to 0$, alors le processus $n^{1/2}(\widehat{F}_\varepsilon(t) - F_\varepsilon(t))$, $t \in \mathbb{R}$, converge en distribution vers un processus gaussien $Z(t)$ de moyenne*

$$\mathbb{E}Z(t) = \mathbb{E}\left[\mathbb{1}(\varepsilon \le t) - F_\varepsilon(t) + \varphi(X, Y, t)\right] = 0,$$

et de fonction covariance

$$\mathrm{Cov}\left(Z(t_1), Z(t_2)\right) = \mathbb{E}\left(\left[\mathbb{1}(\varepsilon \le t_1) - F_\varepsilon(t_1) + \varphi(X, Y, t_1)\right]\left[\mathbb{1}(\varepsilon \le t_2) - F_\varepsilon(t_2) + \varphi(X, Y, t_2)\right]\right).$$

(ii) *Si $a_n = Cn^{-1/4}$, avec $C > 0$, alors le processus $n^{1/2}(\widehat{F}_\varepsilon(t) - F_\varepsilon(t))$, $t \in \mathbb{R}$, converge en distribution vers un processus gaussien $\widetilde{Z}(t)$ de moyenne*

$$\mathbb{E}\widetilde{Z}(t) = \frac{C^2 \mu_K}{2}\int \frac{\partial^2}{\partial x^2}\mathbb{E}\left[\varphi(x, Y, t)|u\right]|_{x=u}\, dF_X(u),$$

et de même fonction de covariance que le processus $Z(t)$.

Le premier point du corrolaire précédent montre que si na_n^4 tend vers 0, alors pour tout $t \in \mathbb{R}$,

$$n^{1/2}(\widehat{F}_\varepsilon(t) - F_\varepsilon(t)) \xrightarrow{d} \mathcal{N}\left(0, VarZ(t)\right). \tag{1.11}$$

De plus, puisque $\mathbb{E}\left[\varphi(X, Y, t)\right] = 0$, un simple calcul montre que

$$\begin{aligned} \mathrm{Var}Z(t) &= \mathbb{E}\left[\mathbb{1}\left(\varepsilon \le t\right) - F_\varepsilon(t) + \varphi\left(X, Y, t\right)\right]^2 \\ &= F_\varepsilon(t)\left(1 - F_\varepsilon(t)\right) + \mathbb{E}\left[\varphi^2(X, Y, t) + 2\mathbb{1}\left(\varepsilon \le t\right)\varphi(X, Y, t)\right]. \end{aligned} \tag{1.12}$$

Mais par le Théorème Central Limite, l'estimateur $F_n(t) = n^{-1}\sum_{i=1}^n \mathbb{1}\left(\varepsilon_i \le t\right)$ de $F_\varepsilon(t)$ basé sur les vrais résidus satisfait

$$n^{1/2}(F_n(t) - F_\varepsilon(t)) \xrightarrow{d} \mathcal{N}\left(0, F_\varepsilon(t)\left(1 - F_\varepsilon(t)\right)\right).$$

Ce résultat, (1.11) et (1.12) montrent que la variance asymptotique obtenue avec l'estimateur $\widehat{F}_\varepsilon(t)$ est inférieure à la variance asymptotique $F_\varepsilon(t)\left(1 - F_\varepsilon(t)\right)$ obtenue avec $F_n(t)$ lorsque

$$\mathbb{E}\left[\varphi^2(X, Y, t) + 2\mathbb{1}\left(\varepsilon \le t\right)\varphi(X, Y, t)\right] \le 0.$$

Dans ce cadre, il ya donc un impact positif causé par l'estimation des résidus sur la loi limite de l'estimateur de $F_\varepsilon(t)$. Notons que ces résultats ne traitent pas le cas où l'ordre de a_n est $n^{-1/5}$, l'ordre optimal de la fenêtre pour l'estimation de $m(\cdot)$.

Dans un article plus récent, Neumeyer et Van Keilegom (2010) ont établi des résultats comparables à ceux obtenus par Akritas et Van Keilegom (2001) dans le cas du modèle de régression hétéroscédastique multiple : $Y = m(X) + \sigma(X)\varepsilon$, $X \in \mathbb{R}^d$, $d \ge 1$.

1.5.2 Estimation adaptative de la densité des résidus

Efromovich (2005, 2007) utilise une méthode adaptative pour estimer la densité f_ε de l'erreur dans le cas des modèles de régression homoscédastique et hétéroscédastique. La méthode est adaptative par rapport à la régularité de f_ε, mesurée par son ordre α de dérivabilité. Un estimateur est alors dit adaptatif s'il ne dépend pas de α mais converge vers f_ε avec la même vitesse que les estimateurs optimaux construits en connaissant α et basés sur les vrais résidus.

Les modèles considérés sont de la forme $Y = m(X) + \varepsilon$ pour le modèle de régression homoscédastique, ou de la forme $Y = m(X) + \sigma(X)\xi$, pour le modèle de régression hétéroscédastique. Ces modèles sont basés sur un échantillon d'observations i.i.d $(X_1, Y_1), \ldots, (X_n, Y_n)$ de même loi que (X, Y). Les variables ξ et ε sont supposées centrées et indépendantes de X. Les fonction $m(\cdot)$ et $\sigma(\cdot)$ sont inconnues et définies sur $[0, 1]$. L'étude d'un estimateur de la densité de l'erreur par Efromovich s'est faite suivant la nature du support de l'erreur. On distinguera le cas où le terme d'erreur est à support borné $[-1, 1]$, et le cas où le terme résiduel est de support non borné $(-\infty, \infty)$. Mais dans cette sous-section, on ne parlera que du dernier cas. Pour le premier cas, le lecteur pourra se référer au papier d'Efromovich (2005).

Dans le cas où le terme d'erreur est de support $(-\infty, \infty)$, l'étude se fait donc avec le modèle de régression homoscédastique $Y = m(X) + \varepsilon$, où la fonction de régression m est supposée inconnue et définie dans $[0, 1]$. Pour estimer la densité f_ε de l'erreur ε, Efromovich utilise un estimateur basé sur un développement en série de cosinus. L'estimation de f_ε nécessite une subdivision des observations en trois sous-échantillons. Le premier sous-échantillon de taille n_1 est utilisé pour estimer la densité marginale p de X. La deuxième partie de l'échantillon (de taille n_1) est réservée à l'estimation de

la fonction de régression m, alors que le dernier sous-échantillon (de taille $n_2 = n - 2n_1$) est réservé à l'estimation de la densité f_ε. On pose, pour tout $u \in [0,1]$,

$$\varphi_0(u) = 1, \quad \varphi_j(u) = \sqrt{2}\cos(\pi j u), \quad j > 0.$$

Les estimateurs de $\widehat{p}$ et $\widehat{m}$ sont alors définis par, pour $x \in [0,1]$,

$$\begin{aligned} \widehat{p}(x) &= \max\left(b_n^{-1}, n_1^{-1}\sum_{\ell=1}^{n_1}\sum_{s=0}^{S}\varphi_s(X_\ell)\varphi_s(x)\right), \\ \widehat{m}(x) &= n_1^{-1}\sum_{\ell=n_1+1}^{2n_1}\sum_{s=0}^{S}\frac{Y_\ell\varphi_s(X_\ell)\varphi_s(x)}{\widehat{p}(X_\ell)}. \end{aligned} \tag{1.13}$$

où $b_n = 4 + \ln\ln(n+20)$, $n_1 = n_1(n)$ désigne le plus petit entier supérieur ou égal à n/b_n, et $S = S_n$ représente le plus petit entier supérieur ou égal à $n^{1/3}$.

Avec l'aide de ces estimateurs de p et m, Efromovich estime les résidus ε_ℓ, $\ell = 2n_1+1, \dots, n$ par

$$\widehat{\varepsilon}_\ell = Y_\ell - \widehat{m}(X_\ell), \quad \ell = 2n_1+1, \dots, n.$$

Pour $t \in \mathbb{R}$, l'estimateur $\widehat{f}_\varepsilon$ de $f_\varepsilon(t)$ est alors défini, suivant la méthode d'estimation de Pinsker (1980), par

$$\widehat{f}_\varepsilon(t) = \sum_{j=0}^{k_n}\widehat{\mu}_j\widehat{\theta}_j\varphi_j(t), \quad \widehat{\theta}_j = (n-2n_1)^{-1}\sum_{\ell=2n_1+1}^{n}\varphi_j\left(\widehat{\varepsilon}_\ell\right),$$

où k_n est le plus petit entier supérieur ou égal à $n^{1/5}b_n$, et les $\widehat{\mu}_j$ sont les estimateurs des coefficients de Fourier $\theta_j = \int_0^1 f_\varepsilon(u)\varphi_j(u)du$. Ces coefficients sont estimés selon la procédure suivante. On subdivise l'ensemble $\mathbb{N}$ des entiers naturels en des blocs non imbriqués B_k, $k = 1, 2, \dots$ et on pose $t_k = 1/\ln(k+2)$. Les $\widehat{\mu}_j$ sont alors définis par

$$\widehat{\mu}_j = \frac{k^{-2}\sum_{s\in B_k}\widehat{\theta}_s^2 - n^{-1}}{k^{-2}\sum_{s\in B_k}\widehat{\theta}_s^2}\mathbb{1}\left(k^{-2}\sum_{s\in B_k}\widehat{\theta}_s^2 > (1+t_k)n^{-1}\right), \quad j \in B_k. \tag{1.14}$$

Pour évaluer la performance de l'estimateur $\widehat{f}_\varepsilon(t)$, Efromovich considère l'estimateur $\overline{f}_\varepsilon(t)$ de f_ε basé sur les vrais résidus. Cet estimateur est défini par

$$\overline{f}_\varepsilon(t) = \sum_{j=0}^{k_n}\overline{\mu}_j\overline{\theta}_j\varphi_j(t), \quad \overline{\theta}_j = (n-2n_1)^{-1}\sum_{\ell=2n_1+1}^{n}\varphi_j\left(\varepsilon_\ell\right),$$

où les coefficients $\overline{\mu}_j$ sont définis comme dans (1.14) en remplaçant seulement les $\widehat{\theta}_j$ par les pseudos-estimateurs $\overline{\theta}_j$ des coefficients θ_j. En définissant l'erreur quadratique moyenne intégrée

$$\mathrm{MISE}(\widehat{f}_\varepsilon, f_\varepsilon) = \mathbb{E}\int_0^1(\widehat{f}_\varepsilon(t) - f_\varepsilon(t))^2dt,$$

Efromovich obtient le résultat suivant.

Theorème 1.6. Efromovich (2005)
On suppose que les fonctions p et m sont de classe C^1 sur $[0,1]$. Alors sous certaines conditions de régularité, on a

$$\mathrm{MISE}(\widehat{f}_\varepsilon, f_\varepsilon) \le \left(1 + \frac{C}{\ln b_n}\right)\mathrm{MISE}(\overline{f}_\varepsilon, f_\varepsilon) + \frac{Cb_n^3}{n},$$

où C est une constante strictement positive.

Dans un article plus récent, Efromovich (2007) montre que le résultat du théorème précédent reste valable sans une procédure de “splitting”(subdivision) des données de l’échantillon.

Dans le cas où la densité f_ε admet une dérivée généralisée d’ordre $\alpha \geq 2$, Efromovich montre que l’estimateur $\overline{f}_\varepsilon$ basé sur les vrais résidus atteint la vitesse de convergence minimax $n^{-2\alpha/(2\alpha+1)}$ pour le risque quadratique moyen intégré. Donc le Théorème 1.6 prouve qu’il n’y a pas de perte (au sens de la vitesse minimax) du fait de ne pas observer les résidus. En conséquence, puisque $\overline{f}_\varepsilon$ est adaptatif par rapport à la régularité de f_ε, il en est de même pour l’estimateur $\widehat{f}_\varepsilon$.

Dans un article récent, Plancade (2008) présente un estimateur nonparamétrique de la densité de l’erreur dans un modèle de régression homoscédastique, basé sur des techniques de sélection de modèle. Avec cette méthode, Plancade propose une majoration du risque quadratique intégré, et obtient la même vitesse minimax que celle obtenue par Efromovich (2005).

1.5.3 Estimation de la fonction variance en régression hétéroscédastique

Dans cette sous-section, nous donnons un exemple sur l’influence de l’estimation la fonction moyenne $m(\cdot)$ sur l’estimation de la fonction variance $V(\cdot)$ dans le cas du modèle de régression hétéroscédastique

$$Y_i = m(x_i) + V^{1/2}(x_i)\varepsilon_i, \quad i = 1, \ldots, n, \tag{1.15}$$

où $x_i = i/n$, et les ε_i sont des variables aléatoires i.i.d, centrées, de variance égale à 1, et admettant des moments d’ordre 4 finis. Dans ce modèle, le paramètre d’intérêt est la fonction V, et on s’intéresse à l’étude de l’impact de m sur l’estimation de V. La qualité de cette estimation est fortement dépendante de la régularité de la fonction de régression m. On souhaite évaluer l’impact de l’estimation de m sur un estimateur de V. Ce problème a été étudié par Wang, Brown, Cai et Levine (2008). Ces auteurs ont montré qu’il est possible d’évaluer explicitement l’impact de m sur l’estimateur de V. Cet impact se mesure à l’aide des erreurs quadratiques moyennes globale et locale définies par

$$R_n = \mathbb{E} \int_0^1 \left(V_n(x) - V(x)\right)^2 dx, \quad R_n(x) = \mathbb{E}\left(V_n(x) - V(x)\right)^2.$$

Ici $V_n(x)$ désigne un estimateur nonparamétrique de $V(x)$. L’estimateur considéré par Wang et al. (2008) est défini comme suit. On considère d’abord un noyau K à support dans $[-1, 1]$. Ensuite, pour $i = 2, \ldots, n-2$, on pose $a_i = (x_i + x_{i-1})/2$ et $b_i = (x_i + x_{i+1})/2$. Enfin, pour $i = 2, \ldots, n-2$, $0 < h < 1/2$ et $x \in [0, 1]$, on définit

$$K_i^h(x) = \int_{a_i}^{b_i} \frac{1}{h} K\left(\frac{x-u}{h}\right) du,$$

et on prend cette intégrale de 0 à $(x_1 + x_2)/2$ pour $i = 1$, et de $(x_{n-1} + x_{n-2})/2$ à 1 pour $i = n-1$. Sous certaines hypothèses sur le noyaux K, on peut vérifier que pour tout $x \in [0, 1]$, $\sum_{i=1}^{n-1} K_i^h(x) = 1$. L’estimateur $V_n(x)$ de $V(x)$ est alors défini par

$$V_n(x) = \frac{1}{2} \sum_{i=1}^{n-1} K_i^h(x) \left(Y_i - Y_{i+1}\right)^2. \tag{1.16}$$

Pour $\alpha > 0$ et $M > 0$, considérons la classe de fonctions M-lipschitziennes

$$\mathcal{L}^\alpha(M) = \left\{ g : \forall\, x, y \in [0,1], \forall\, k = 0, \ldots, \lfloor\alpha\rfloor - 1,\ |g^{(k)}| \leq M,\ \left|g^{(\lfloor\alpha\rfloor)}(x) - g^{(\lfloor\alpha\rfloor)}(y)\right| \leq M|x-y|^{\alpha'} \right\},$$

où $\lfloor\alpha\rfloor$ est le plus grand entier naturel inférieur à α, et $\alpha' = \alpha - \lfloor\alpha\rfloor$. On a alors le résultat suivant.

Theorème 1.7. Wang, Brown, Cai et Levine (2008)
On considère le modèle de régression (1.15), où $x_i = i/n$, et les ε_i sont des variables aléatoires i.i.d, centrées, de variance égale à 1, *et admettant des moments d'ordre* 4 *finis. On suppose qu'il existe des constantes strictement positives α, β, M_1 et M_2 telles que $m \in \mathcal{L}^\alpha(M_1)$ et $V \in \mathcal{L}^\beta(M_2)$. Alors sous des hypothèses convenables, la fenêtre optimale h_n pour l'estimateur $V_n(x)$ de $V(x)$ est de l'ordre de $n^{-1/(1+2\beta)}$. De plus, pour un tel choix optimal de h_n, la vitesse de convergence mimimax pour les quantités R_n et $R_n(x)$ est de l'ordre de* $\max\{n^{-4\alpha}, n^{-2\beta/(2\beta+1)}\}$.

A l'aide de ce thèorème, on peut comparer la performance (en terme de vitesse minimax) de l'estimateur $V_n(x)$ à celle de l'estimateur $\widehat{V}_n(x)$ basé sur l'estimation de m par $\widehat{m}_n$. Cet estimateur $\widehat{V}_n(x)$ est de la forme

$$\widehat{V}_n(x) = \sum_{i=1}^{n-1} w_i(x) \left(Y_i - \widehat{m}_n(x_i)\right)^2, \tag{1.17}$$

où les $w_i(x)$ sont des fonctions poids. On note qu'avec l'estimateur $\widehat{V}_n(x)$, la vitesse de convergence minimax $\max\{n^{-4\alpha}, n^{-2\beta/(2\beta+1)}\}$ ne peut être obtenue que si la fonction moyenne m est estimée par un estimateur de $\widehat{m}_n$ faiblement biaisé. C'est ce qui a incité Brown, Cai et Levine (2008) à prendre un estimateur $\widehat{m}_n$ de m tel que $\widehat{m}_n(x_i) = Y_{i+1}$. Ce qui, reporté dans (1.17), conduit à un estimateur du type (1.16). Un tel estimateur a une variance assez élevée et un biais suffisamment petit, pour n suffisamment grand. Mais les auteurs ont prouvé qu'une grande variance de $\widehat{m}_n$ ne peut pas affecter la vitesse de convergence de $\widehat{V}_n$. Donc finalement, pour l'estimation de la fonction V, un estimateur optimal $\widehat{m}_n$ est celui de biais minimum, et non nécessairement celui d'erreur quadratique mimimale. Un enseignement important est que le carré du biais de $\widehat{m}_n$ joue un rôle plus important que sa variance. En conséquence, utiliser un estimateur qui serait optimal pour l'estimation de m n'est pas intéressant ici, car un tel estimateur égalise asymptotiquement le carré du biais et la variance.

1.5.4 Estimation de la densité des résidus basée sur un estimateur de Nadaraya-Watson de la fonction de régression

Le problème de l'estimation nonparamétrique de la densité f des résidus a été considéré par Cheng (2005) dans le cadre du modèle de régression nonparamétrique $Y = m(X) + \varepsilon$. Dans ce modèle, la fonction de régression m est définie sur $[0, 1]$, et les estimateurs proposés se construisent en utilisant les observations $(X_1, Y_1), \ldots, (X_n, Y_n)$. Ces observations sont scindées en deux parties. La première partie est destinée à l'estimation des résidus $\varepsilon_i = Y_i - m(X_i)$, tandis que la seconde partie des observations est réservée à la construction de l'estimateur de f. Les estimateurs $\widehat{\varepsilon}_i$ des résidus ε_i s'obtiennent à partir des estimations des quantités $m(X_i)$. Pour ce faire, Cheng considère un entier r_n dépendant de n, et satisfaisant

$$0 < r_n \leq n/2, \quad \lim_{n\to\infty} r_n = \infty, \quad \lim_{n\to\infty} (n - r_n) = \infty.$$

Il utilise les r_n premières observations $(X_1, Y_1), \ldots, (X_{r_n}, Y_{r_n})$ pour construire l'estimateur de la fonction $m(x)$. Cet estimateur de $m(x)$ est celui de Nadaraya-Watson basé sur les données $(X_1, Y_1), \ldots, (X_{r_n}, Y_{r_n})$:

$$m_n(x) = \frac{\sum_{i=1}^{r_n} Y_i K\left(\frac{X_i - x}{h_n}\right)}{\sum_{i=1}^{r_n} K\left(\frac{X_i - x}{h_n}\right)}, \quad x \in [0, 1],$$

où h_n est une fenêtre strictement positive tendant vers 0 quand n tend vers l'infini, et K une fonction intégrable sur $\mathbb{R}$ et d'intégrale 1.

Le reste des observations $(X_{r_n+1}, Y_{r_n+1}), \ldots, (X_n, Y_n)$ est utilisé pour estimer les résidus ε_i par

$$\widehat{\varepsilon}_i = Y_i - m_n(X_i), \quad r_n + 1 \leq i \leq n.$$

L'estimateur nonparamétrique de la densité des résidus construit par Cheng est alors défini par

$$\widehat{f}_n(t) = \frac{1}{2(n - r_n)a_n} \sum_{i=r_n+1}^{n} \mathbb{1}\left(t - a_n < \widehat{\varepsilon}_i \leq t + a_n\right), \quad t \in \mathbb{R}.$$

Avec cet estimateur, Cheng (2005) obtient le résultat suivant.

Theorème 1.8. *Soit $t \in [0, 1]$ tel que $f(t) > 0$. Supposons que $0 \leq r_n \leq n/2$ tel que*

$$\lim_{n\to\infty} (n - r_n)a_n^3 = 0, \quad \lim_{n\to\infty} (n - r_n)a_n = \infty, \quad \lim_{n\to\infty} \frac{(n - r_n)a_n \log r_n}{r_n h_n} = 0. \tag{1.18}$$

On suppose également que la densité g des X_i est localement lipchitzienne sur $[0, 1]$. Alors sous d'autres hypothèses de régularité, on a la convergence en distribution suivante :

$$\sqrt{2(n - r_n)a_n} \left(\frac{\widehat{f}_n(t) - f(t)}{\sqrt{f(t)}} \right) \xrightarrow{d} N(0, 1),$$

où $N(0, 1)$ désigne la loi normale centrée réduite.

Il a été démontré dans la littérature statistique que $n^{-2/5}$ est la vitesse optimale de convergence obtenue avec l'estimation nonparamétrique de la densité d'une variable aléatoire réelle ζ à partir d'un échantillon de variables aléatoires i.i.d $\zeta_1, \zeta_2, \ldots, \zeta_n$. Pour ce résultat, on peut, par exemple, se référer aux ouvrages de Bosq et Lecoutre (1987), Scott (1992), Wand et Jones (1995). Mais pour $0 \leq r_n \leq n/2$, le résultat du théorème précédent montre que la vitesse $n^{-2/5}$ pour l'estimateur $\widehat{f}_n(t)$ ne peut-être atteinte que si la fenêtre a_n est d'ordre $n^{-1/5}$. Mais pour un tel ordre, la première condition dans (1.18) ne peut pas être satisfaite. Donc sous les conditions du thèorème précédent, l'estimateur $\widehat{f}_n(t)$ ne peut pas atteindre la vitesse optimale $n^{-2/5}$, ni même s'en approcher. En effet, (1.18) implique que $a_n = o\left(1/n^{1/3}\right)$, et que la vitesse de convergence de $\widehat{f}_n(t)$ est $o\left(1/n^{1/3}\right)$.

Cette thèse améliore les résultats de Cheng (2005). En effet, nous verrons que sous des hypothèses convenables, les estimateurs que nous proposerons pour estimater la loi f des résidus pourront atteindre la vitesse de convergence $n^{-2/5}$ pour $\dim(X) \leq 2$, où $\dim(X)$ désigne la dimension de la variable explicative X.

Chapitre 2

Contribution de la thèse

2.1 Introduction

La revue de la littérature faite au Chapitre 1 montre que la plupart des auteurs cités précédemment ont utilisé les résidus estimés pour construire un estimateur d'une distribution de l'erreur. Mais aucun d'entre eux ne s'est attaché à étudier l'impact de la dimension de la variable explicative sur l'estimateur de la loi f des erreurs, ni d'évaluer l'influence de la fenêtre de première étape (utilisée pour estimer la fonction de régression) sur l'estimateur final de la densité des résidus. La thèse s'attachera donc à évaluer l'impact de la dimension de la variable X sur l'estimation de la densité f. Nous tenterons également de déterminer les vitesses de convergence ponctuelle des estimateurs nonparamétriques de f. Un de nos objectifs majeurs sera aussi de caractériser les façons optimales de choisir les fenêtres de première et deuxième étapes utilisées pour estimer f.

Nous donnons maintenant une briève présentation de nos résulats qui seront établis dans les deux prochains chapitres de la thèse.

2.2 Estimateur conditionnel nonparamétrique de la densité des résidus

Pour mieux illustrer l'effet de la dimension de la variable explicative X sur l'estimation de la densité f des résidus du modèle de régression (1.1), nous considérons d'abord une méthode naïve d'estimation de f basée sur la relation

$$f(\epsilon|x) = \varphi\left(m(x) + \epsilon|x\right),$$

où $f(\cdot|x)$ et $\varphi(\cdot|x)$ désignent respectivement les densités de ε et Y sachant que $X = x$. En utilisant l'indépendance de X et ε, on a donc

$$f(\epsilon) = f(\epsilon|x) = \varphi\left(m(x) + \epsilon|x\right).$$

Suivant cette idée, on peut donc déduire un estimateur de $f(\epsilon)$ à partir d'une estimation de $\varphi(y|x)$ et de $m(x)$. Par conséquent, un estimateur $\widetilde{f}_n(\epsilon|x)$ de $f(\epsilon)$ est défini par

$$\widetilde{f}_n(\epsilon|x) = \frac{\frac{1}{nh_0^d h_1}\sum_{i=1}^n K_0\left(\frac{X_i - x}{h_0}\right) K_1\left(\frac{Y_i - \widehat{m}_n(x) - \epsilon}{h_1}\right)}{\frac{1}{nh_0^d}\sum_{i=1}^n K_0\left(\frac{X_i - x}{h_0}\right)},$$

où h_0, h_1 et b_1 désignent des fenêtres positives, K_0 et K_1 sont des fonctions noyaux définies respectivement sur $\mathbb{R}^d$ et $\mathbb{R}$, et $\widehat{m}_n(x)$ l'estimateur de Nadaraya-Watson (1964) de $m(x)$ defini par

$$\widehat{m}_n(x) = \frac{\sum_{j=1}^n Y_j K_0\left(\frac{X_j - x}{b_0}\right)}{\sum_{j=1}^n K_0\left(\frac{X_j - x}{b_0}\right)},$$

où b_0 est une fenêtre positive. Le théorème suivant, qui sera démontré dans la suite de cette thèse, permet de mieux illustrer l'effet négatif de la dimension de X sur le comportement asymptotique de l'estimateur $\widetilde{f}_n(\epsilon|x)$.

Theorème 2.1. *Considérons*

$$\mu_1(x,\epsilon) = \frac{\partial^2 \varphi(x, m(x)+\epsilon)}{\partial^2 x}\int z K_0(z) z^\top dz, \quad \mu_2(x,\epsilon) = \frac{\partial^2 \varphi(x, m(x)+\epsilon)}{\partial^2 y}\int v^2 K_1(v)dv,$$

et supposons que b_0, h_0 et h_1 décroissent vers 0 et satisfont $nh_0^{2d}/\ln n \to \infty$, $\ln(1/h_0)/\ln(\ln n) \to \infty$ et

$$nh_0^d h_1 \to \infty, \quad \left(\frac{nh_0^d}{h_1}\right)\left(b_0^4 + \frac{\ln n}{nb_0^d}\right) = o(1),$$

lorsque $n \to \infty$. Alors sous des conditions de régularité sur m, g, φ, K_0 and K_1, on a

$$\sqrt{nh_0^d h_1}\left(\widetilde{f}_n(\epsilon|x) - \overline{\widetilde{f}}_n(\epsilon|x)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{f(\epsilon|x)}{g(x)}\int\int K_0^2(z)K_1^2(v)dzdv\right),$$

où

$$\overline{\widetilde{f}}_n(\epsilon|x) = f(\epsilon|x) + \frac{h_0^2\mu_1(x,\epsilon)}{2g(x)} + \frac{h_1^2\mu_2(x,\epsilon)}{2g(x)} + o\left(h_0^2 + h_1^2\right).$$

Le résultat de ce théorème suggère que pour la normalité asymptotique de l'estimateur $\widetilde{f}_n(\epsilon|x)$, les fenêtres optimales h_0 et h_1 sont celles qui minimisent le développement quadratique moyenne asymptotique

$$AMSE\left(\widetilde{f}_n(\epsilon|x)\right) = \left[\frac{h_0^2\mu_1(x,\epsilon)}{2g(x)} + \frac{h_1^2\mu_2(x,\epsilon)}{2g(x)}\right]^2 + \frac{f(\epsilon|x)\int K_0^2(z)dz\int K_1^2(v)dv}{nh_0^d h_1 g(x)}.$$

Un simple calcul montre que les fenêtres optimales h_0 et h_1 sont toutes de l'ordre de $n^{-1/(d+5)}$, conduisant à une vitesse de convergence optimale $n^{-2/(d+5)}$ pour l'estimateur $\widetilde{f}_n(\epsilon|x)$. Par conséquent, dans le cas où $d = 1$, cette vitesse de convergence est de l'ordre de $n^{-2/3}$, ce qui est pire que la vitesse optimale $n^{-2/5}$ atteinte dans le cadre de l'estimation d'une densité univariée. Pour la vitesse optimale de l'estimateur d'une densité univariée, on pourra consulter, par exemple, les ouvrages de

Bosq and Lecoutre (1987), Scott (1992), Wand and Jones (1995). On note également que l'exposant $2/(d+5)$ décroît vers 0 lorsque d devient de plus en plus grand. Cette situation illustre donc l'impact négatif de la dimension de X sur la performance (au sens de la vitesse de convergence optimale) de l'estimateur $\widetilde{f}_n(\epsilon|x)$. C'est le problème du "fléau de la dimension". Ce problème est dû au conditionnement par x dans l'expression $f(\epsilon) = f(\epsilon|x) = \varphi(m(x) + \epsilon|x)$, où l'on identifie la densité non conditionnelle $f(\epsilon)$ à la densité conditionnelle $f(\epsilon|x)$ sous l'hypothèse d'indépendance de ε et X. Il convient également d'ajouter que si on voulait utiliser l'estimateur $\widetilde{f}_n(\epsilon|x)$, il faudrait résoudre le problème du choix de x. En effet, même si la densité $f(\epsilon)$ ne dépend pas de x, l'estimateur $\widetilde{f}_n(\epsilon|x)$ en dépend.

Pour palier ce problème du "fléau de la dimension", il faut donc "déconditionner"dans l'expression ci-dessus de $f(\epsilon)$. Deux approches sont alors proposées dans la suite cette thèse. Ces approches sont résumées dans les deux sections suivantes.

2.3 Estimation de la densité de l'erreur par utilisation des résidus estimés

Cette première approche consiste, dans un premier temps, à estimer nonparamétriquement les résidus ε_i du modèle (1.1) par

$$\widehat{\varepsilon}_i = Y_i - \widehat{m}_{in}, \quad i = 1, \dots, n,$$

où $\widehat{m}_{in} = \widehat{m}_{in}(X_i)$ désigne le "leave-one out"estimateur à noyau de $m(X_i)$ défini par

$$\widehat{m}_{in} = \frac{\sum_{\substack{j=1 \\ j\neq i}}^{n} Y_j K_0\left(\frac{X_i - X_j}{b_0}\right)}{\sum_{\substack{j=1 \\ j\neq i}}^{n} K_0\left(\frac{X_i - X_j}{b_0}\right)}.$$

Dans un deuxième temps, on utilise ces résidus estimés, comme si c'était les vrais, pour construire un estimateur nonparamétrique de $f(\epsilon)$. Cette construction tient compte du fait que les $\widehat{m}_n(X_i)$ peuvent être des estimateurs biaisés des $m(X_i)$ lorsque les variables X_i sont très proches des bords de leur support $\mathcal{X}$. Par conséquent, l'estimateur de $f(\epsilon)$ est construit en prenant les observations X_i dans un ensemble ouvert $\mathcal{X}_0$ intérieur à $\mathcal{X}$. L'estimateur de $f(\epsilon)$ est donc défini par

$$\widehat{f}_{1n}(\epsilon) = \frac{1}{b_1 \sum_{i=1}^{n} \mathbb{1}\left(X_i \in \mathcal{X}_0\right)} \sum_{i=1}^{n} \mathbb{1}\left(X_i \in \mathcal{X}_0\right) K_1\left(\frac{\widehat{\varepsilon}_i - \epsilon}{b_1}\right).$$

En principe, on peut supposer que $\mathcal{X}_0$ est suffisamment proche de $\mathcal{X}$ de telle sorte que $\widehat{f}_{1n}(\epsilon)$ se rapproche considérablement de l'estimateur "classique"$\sum_{i=1}^{n} K\left((\widehat{\varepsilon}_i - \epsilon)/b_1\right)/(nb_1)$. Néanmoins, dans la suite de cette thèse, nous considérerons un sous-ensemble fixé $\mathcal{X}_0$, pour des raisons de commodité. Notons aussi que l'estimateur $\widehat{f}_{1n}(\epsilon)$ ne dépend d'aucun paramètre inconnu, comme désiré dans la pratique. Ceci contraste avec l'estimateur idéal nonparamétrique

$$\widetilde{f}_{1n}(\epsilon) = \frac{1}{b_1 \sum_{i=1}^{n} \mathbb{1}\left(X_i \in \mathcal{X}_0\right)} \sum_{i=1}^{n} \mathbb{1}\left(X_i \in \mathcal{X}_0\right) K_1\left(\frac{\varepsilon_i - \epsilon}{b_1}\right),$$

qui dépend en particulier des résidus non observés ε_i. Cet estimateur $\widetilde{f}_{1n}(\epsilon)$ est très proche de l'estimateur $\widehat{f}_{1n}(\epsilon)$, comme le suggère le théorème suivant.

Theorème 2.2. *Supposons que b_0 and b_1 décroissent vers 0 telles que $\ln(1/b_0)/\ln(\ln n) \to \infty$, $nb_0^{d^*}/\ln n \to \infty$, $d^* = \sup\{d+2, 2d\}$, et $n^{(d+8)}b_1^{7(d+4)} \to \infty$ lorsque $n \to \infty$. Alors sous certaines conditions de régularité sur m, g, f, K_0 et K_1, on a*

$$\widehat{f}_{1n}(\epsilon) - \widetilde{f}_{1n}(\epsilon) = O_{\mathbb{P}}\Big(R_n(b_0,b_1)\Big)^{1/2}, \quad \widehat{f}_{1n}(\epsilon) - f(\epsilon) = O_{\mathbb{P}}\Big(AMSE(b_1) + R_n(b_0,b_1)\Big)^{1/2},$$

où

$$AMSE(b_1) = \mathbb{E}_n\left[\left(\widetilde{f}_{1n}(\epsilon) - f(\epsilon)\right)^2\right] = O_{\mathbb{P}}\left(b_1^4 + \frac{1}{nb_1}\right),$$

et

$$R_n(b_0,b_1) = b_0^4 + \left[\frac{1}{(nb_1^5)^{1/2}} + \left(\frac{b_0^d}{b_1^3}\right)^{1/2}\right]^2\left(b_0^4 + \frac{1}{nb_0^d}\right)^2 + \left[\frac{1}{b_1} + \left(\frac{b_0^d}{b_1^7}\right)^{1/2}\right]^2\left(b_0^4 + \frac{1}{nb_0^d}\right)^3.$$

Les résultats de ce théorème donnent une première idée de l'impact de l'estimation des résidus sur l'estimateur nonparamétrique de la densité $f(\epsilon)$.

Le théorème suivant détermine la façon optimale de choisir la fenêtre de première étape b_0. A notre connaissance, cet aspect n'a pas encore été étudié dans la littérature statistique. Dans ce qui suit, $a_n \asymp b_n$ signifie que $a_n = O(b_n)$ et $b_n = O(a_n)$, c'est à dire il existe une constante $C > 0$ telle que $|a_n|/C \leq |b_n| \leq C|a_n|$, pour n suffisamment grand.

Theorème 2.3. *On considère la fenêtre*

$$b_0^* = b_0^*(b_1) = \arg\min_{b_0} R_n(b_0,b_1),$$

où la minimisation se fait sur l'ensemble des fenêtres b_0 satisfaisant les condtions du théorème précédent. Alors la fenêtre b_0^* vérifie

$$b_0^* \asymp \max\left\{\left(\frac{1}{n^2b_1^3}\right)^{\frac{1}{d+4}}, \left(\frac{1}{n^3b_1^7}\right)^{\frac{1}{2d+4}}\right\},$$

et on a

$$R_n(b_0^*,b_1) \asymp \max\left\{\left(\frac{1}{n^2b_1^3}\right)^{\frac{4}{d+4}}, \left(\frac{1}{n^3b_1^7}\right)^{\frac{4}{2d+4}}\right\}.$$

De ce théorème, on déduit le résultat suivant qui donne les conditions pour lesquelles l'estimateur $\widehat{f}_{1n}(\epsilon)$ atteind la vitesse optimale $n^{-2/5}$ lorsque $b_0 = b_0^*$.

Theorème 2.4. *On considère la fenêtre*

$$b_1^* = \arg\min_{b_1}\Big(AMSE(b_1) + R_n(b_0^*,b_1)\Big),$$

où $b_0^* = b_0^*(b_1)$ est definie comme dans le théorème précédent. Alors

1. *Pour $d \leq 2$, la fenêtre b_1^* satisfait*
$$b_1^* \asymp \left(\frac{1}{n}\right)^{\frac{1}{5}},$$
et on a
$$\left(AMSE(b_1^*) + R_n(b_0^*, b_1^*)\right)^{\frac{1}{2}} \asymp \left(\frac{1}{n}\right)^{\frac{2}{5}}.$$

2. *Pour $d \geq 3$, b_1^* satisfait*
$$b_1^* \asymp \left(\frac{1}{n}\right)^{\frac{3}{2d+11}},$$
et on a
$$\left(AMSE(b_1^*) + R_n(b_0^*, b_1^*)\right)^{\frac{1}{2}} \asymp \left(\frac{1}{n}\right)^{\frac{6}{2d+11}}.$$

Ces résultats montrent que pour $d \leq 2$, la vitesse de convergence de la différence $\widehat{f}_{1n}(\epsilon) - f(\epsilon)$ est d'ordre $n^{-2/5}$, ce qui correspond à la vitesse de convergence optimale dans le cas de l'estimation de la densité d'une variable univariée. Donc dans ce cas, il ya un impact positif de l'estimation des résidus sur l'estimateur de $f(\epsilon)$. Mais pour $d \geq 3$, la vitesse le taux de convergence $n^{-2/5}$ ne peut pas être atteinte avec l'estimateur $\widehat{f}_{1n}(\epsilon)$.

Nous obtenons également le résulat de normalité asymptotique suivant.

Theorème 2.5. *Supposons que*
$$nb_0^{d+4} = O(1), \quad nb_0^4 b_1 = o(1), \quad nb_0^d b_1^3 \to \infty,$$
lorsque n tend vers ∞. Alors sous des conditions de régularité, on a
$$\sqrt{nb_1}\left(\widehat{f}_{1n}(\epsilon) - \overline{f}_{1n}(\epsilon)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{f(\epsilon)}{\mathbb{P}\left(X \in \mathcal{X}_0\right)} \int K_1^2(v)dv\right),$$
où
$$\overline{f}_{1n}(\epsilon) = f(\epsilon) + \frac{b_1^2}{2} f^{(2)}(\epsilon) \int v^2 K_1(v)dv + o\left(b_1^2\right).$$

La deuxième approche utilisée pour l'estimation de la densité f est résumée dans la sous-section suivante.

2.4 Estimation de la densité de l'erreur par intégration d'une loi conditionnelle

Cette approche consiste d'abord à remarquer que
$$f(\epsilon) = \int \varphi\left(\epsilon + m(x)|x\right) g(x)dx = \int \varphi\left(x, \epsilon + m(x)\right) dx,$$
où g désigne la densité marginale de X, et $\varphi(\cdot,\cdot)$ la densité conjointe du couple (X, Y). Cette formule suggère donc d'estimer, dans un second temps, $f(\epsilon)$ par
$$\widehat{f}_{2n}(\epsilon) = \int \widehat{\varphi}_n\left(x, \epsilon + \widehat{m}_n(x)\right) dx,$$

où $\widehat{m}_n(x)$ désigne l'estimateur à noyau de Nadaraya-Watson (1964) de $m(x)$, et $\widehat{\varphi}_n$ l'estimateur nonparamétrique de φ. Ces estimateurs sont définis comme suit. On considère des fenêtres $b_0 = b_0(n)$ et $b_1 = b_1(n)$ associées à la variable X, et une fenêtre $h = h(n)$ associé à la variable Y. On suppose que K_0 et K_1 sont des fonctions noyaux définis dans $\mathbb{R}^d$, et que K_2 désigne une fonction noyau défini dans $\mathbb{R}$. Pour tout $(x, y) \in \mathbb{R}^d \times \mathbb{R}$, les estimateurs $\widehat{m}_n(x)$ et $\widehat{\varphi}_n(x, y)$ sont définis par

$$\begin{aligned} \widehat{m}_n(x) &= \frac{\sum_{j=1}^n Y_j K_0\left(\frac{X_j - x}{b_0}\right)}{\sum_{j=1}^n K_0\left(\frac{X_j - x}{b_0}\right)}, \\ \widehat{\varphi}_n(x, y) &= \frac{1}{n b_1^d h} \sum_{i=1}^n K_1\left(\frac{X_i - x}{b_1}\right) K_2\left(\frac{Y_i - y}{h}\right). \end{aligned}$$

On considère également

$$\widetilde{f}_{2n}(\epsilon) = \int \widehat{\varphi}_n(x, \epsilon + m(x))\, dx,$$

l'estimateur par de f basé sur la fonction de régression m. Avec l'aide de ces estimateurs, on obtient d'abord le théorème suivant.

Theorème 2.6. *On suppose que b_0, b_1 et h décroissent vers 0 telles que $\ln(1/b_0)/\ln(\ln n) \to \infty$, $b_0^d/(n b_0^{2d})^p = O(b_0^{2p})$, $p \in [0, 6]$, $n b_1^{2d} \to \infty$ et $n^{(d+8)} h^{7(d+4)} \to \infty$ lorsque $n \to \infty$. Alors, sous des conditions de régularité sur g, m, f, φ, et K_j, $j = 0, 1, 2$, on a*

$$\widehat{f}_{2n}(\epsilon) - f(\epsilon) = O_{\mathbb{P}}\Big(AMSE(b_1, h) + RT_n(b_0, b_1, h)\Big)^{1/2},$$

où

$$AMSE(b_1, h) = \mathbb{E}_n\left[\left(\widetilde{f}_{2n}(\epsilon) - f(\epsilon)\right)^2\right] = O_{\mathbb{P}}\left(b_1^4 + h^4 + \frac{1}{n b_1}\right),$$

et

$$\begin{aligned} RT_n(b_0, b_1, h) &= b_0^4 + \left(b_0^d \vee b_1^d\right)\left[\frac{1}{n b_1^d h^3}\left(b_0^4 + \frac{1}{n b_0^d}\right) + \frac{1}{n b_0^d}\right] \\ &\quad + \left(b_0^d \vee b_1^d\right)\left[\frac{1}{n b_1^d h^5}\left(b_0^4 + \frac{1}{n b_0^d}\right)^2 + \frac{1}{n^2 b_0^{2d} h^3}\right] \\ &\quad + \frac{1}{h^2}\left(b_0^4 + \frac{1}{n b_0^d}\right)^3 + \frac{b_0^d \vee b_1^d}{h^7}\left(b_0^4 + \frac{1}{n b_0^d}\right)^3. \end{aligned}$$

En se basant sur ce théorème, on retrouve des résultats similaires à ceux obtenus avec l'estimateur $\widehat{f}_{1n}(\epsilon)$, notamment ceux relatifs aux choix optimaux des fenêtres de première et deuxième étape pour l'estimation de $f(\epsilon)$.

- Choix optimal de la fenêtre b_0

Theorème 2.7. *On pose $b_0 = b_1$, puis on considère la fenêtre*

$$b_0^* = b_0^*(h) = \arg\min_{b_0} RT_n(b_0, b_0, h),$$

où la minimisation se fait sur l'ensemble des fenêtres b_0 satisfaisant les hypothèses du théorème précédent. Alors b_0^* vérifie

$$b_0^* \asymp \max\left\{\left(\frac{1}{n^2h^3}\right)^{\frac{1}{d+4}}, \left(\frac{1}{n^3h^7}\right)^{\frac{1}{2d+4}}\right\},$$

et on a

$$RT_n(b_0^*, b_0^*, h) \asymp \frac{1}{n} + \max\left\{\left(\frac{1}{n^2h^3}\right)^{\frac{4}{d+4}}, \left(\frac{1}{n^3h^7}\right)^{\frac{4}{2d+4}}\right\}.$$

• Choix optimal de la fenêtre h

Theorème 2.8. *On considère la fenêtre*

$$h^* = \arg\min_h \Big(AMSE(b_0^*, h) + RT_n(b_0^*, b_0^*, h) \Big),$$

où $b_0^* = b_0^*(h)$ est définie comme dans le théorème précédent. Alors

1. *Pour $d \leq 2$, la fenêtre h^* vérifie*
$$h^* \asymp \left(\frac{1}{n}\right)^{\frac{1}{5}},$$
et on a
$$\Big(AMSE(b_0^*, h^*) + RT_n\left(b_0^*, h^*, h^*\right)\Big)^{\frac{1}{2}} \asymp \left(\frac{1}{n}\right)^{\frac{2}{5}}.$$

2. *Pour $d \geq 3$, h^* satisfait*
$$h^* \asymp \left(\frac{1}{n}\right)^{\frac{3}{2d+11}},$$
et on a
$$\Big(AMSE(b_0^*, h^*) + RT_n(b_0^*, b_0^*, h^*)\Big)^{\frac{1}{2}} \asymp \left(\frac{1}{n}\right)^{\frac{6}{2d+11}}.$$

La conclusion des résulats de ce théorème est la même que celle du théorème similaire obtenu avec l'estimateur $\widehat{f}_{1n}(\epsilon)$.

• Normalité asymptotique

Theorème 2.9. *Supposons que*

$$nb_0^{d+4} = O(1), \quad nb_0^4 h = o(1), \quad nb_0^d h^3 \to \infty,$$

lorsque $n \to \infty$. Alors sous certaines conditions de régularité on a,

$$\sqrt{nh}\left(\widehat{f}_{2n}(\epsilon) - \overline{f}_{2n}(\epsilon)\right) \xrightarrow{d} \mathcal{N}\left(0, f(\epsilon)\int K_2^2(v)dv\right),$$

avec

$$\begin{aligned}\overline{f}_{2n}(\epsilon) &= f(\epsilon) + \frac{b_0^2}{2}\int \mathbb{1}\left(x \in \mathcal{X}\right) \frac{\partial^2 \varphi(x, \epsilon + m(x))}{\partial^2 x} dx \int z K_1(z) z^\top dz \\ &\quad + \frac{h^2}{2}\int \mathbb{1}\left(x \in \mathcal{X}\right) \frac{\partial^2 \varphi(x, \epsilon + m(x))}{\partial^2 y} dx \int v^2 K_2(v)dv + o\left(b_0^2 + h^2\right).\end{aligned}$$

Pour finir la thèse, nous réaliserons des simulations numériques pour valider et mieux mettre en exergue les résultats obtenus avec les estimateurs $\widehat{f}_{1n}$ et $\widehat{f}_{2n}$. Nous comparerons les performances de ces estimateurs en terme d'erreurs quadratiques moyennes globales et locales. Nous présenterons également des perspectives de recherche pour nos futurs travaux.

Chapitre 3

Nonparametric kernel estimation of the probability density function of regression errors using estimated residuals

Abstract : In this chapter we deal with the nonparametric density estimation of the regression error term assuming its independence with the covariate. The difference between the feasible estimator which uses the estimated residuals and the unfeasible one using the true residuals is studied. An optimal choice of the bandwidth used to estimate the residuals is given. We also study the asymptotic normality of the feasible kernel estimator and its rate-optimality.

3.1 Introduction

Consider a sample $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$ of independent and identically distributed (i.i.d) random variables, where Y is the univariate dependent variable and the covariate X is of dimension d. Let $m(\cdot)$ be the conditional expectation of Y given X and let ε be the related regression error term, so that the regression error model is

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, \ldots, n. \tag{3.1.1}$$

We wish to estimate the probability distribution function (p.d.f) of the regression error term, $f(\cdot)$, using the nonparametric residuals. Our potential applications are as follows. First, an estimation of the p.d.f of ε is an important tool for understanding the residuals behavior and therefore the fit of the regression model (3.1.1). This estimation of $f(\cdot)$ can be used for goodness-of-fit tests of a specified error distribution in a parametric regression setting. Some examples can be founded in Loynes (1980), Akritas and Van Keilegom (2001), Cheng and Sun (2008). The estimation of the

density of the regression error term can also be useful for testing the symmetry of the residuals. See Ahmad et Li (1997), Dette et *al.* (2002). Another interest of the estimation of f is that it can be used for constructing nonparametric estimators for the density and hazard function of Y given X, as related in Van Keilegom and Veraverbeke (2002). This estimation of f is also important when are interested in the estimation of the p.d.f of the response variable Y. See Escanciano and Jacho-Chavez (2010). Note also that an estimation of the p.d.f of the regression errors can be useful for proposing a mode forecast of Y given $X = x$. This mode forecast is based on an estimation of $m(x) + \arg\min_{\epsilon \in \mathbb{R}} f(\epsilon)$.

Relatively little is known about the nonparametric estimation of the p.d.f and the cumulative distribution function (c.d.f) of the regression error. Up to few exceptions, the nonparametric literature focuses on studying the distribution of Y given X. See Roussas (1967, 1991), Youndjé (1996) and references therein. Akritas and Van Keilegom (2001) estimate the cumulative distribution function of the regression error in heteroscedastic model. The estimator proposed by these authors is based on a nonparametric estimation of the residuals. Their result show the impact of the estimation of the residuals on the limit distribution of the underlying estimator of the cumulative distribution function. Müller, Schick and Wefelmeyer (2004) consider the estimation of moments of the regression error. Quite surprisingly, under appropriate conditions, the estimator based on the true errors is less efficient than the estimator which uses the nonparametric estimated residuals. The reason is that the latter estimator better uses the fact that the regression error ε has mean zero. Efromovich (2005) consider adaptive estimation of the p.d.f of the regression error. He gives a nonparametric estimator based on the estimated residuals, for which the Mean Integrated Squared Error (MISE) attains the minimax rate. Fu and Yang (2008) study the asymptotic normality of the estimators of the regression error p.d.f in nonlinear autoregressive models. Cheng (2005) establishes the asymptotic normality of an estimator of $f(\cdot)$ based on the estimated residuals. This estimator is constructed by splitting the sample into two parts : the first part is used for the construction of estimator of $f(\cdot)$, while the second part of the sample is used for the estimation of the residuals.

The focus of this chapter is to estimate the p.d.f of the regression error using the estimated residuals, under the assumption that the covariate X and the regression error ε are independent. In a such setup, it would be unwise to use a conditional approach based on the fact that $f(\epsilon) = f(\epsilon|x) = \varphi(m(x) + \epsilon|x)$, where $\varphi(\cdot|x)$ is the p.d.f of Y given $X = x$. Indeed, the estimation of $m(\cdot)$ and $\varphi(\cdot|x)$ are affected by the curse of dimensionality, so that the resulting estimator of $f(\cdot)$ would have considerably a slow rate of convergence if the dimension of X is high. The approach proposed here uses a two-steps procedure which, in a first step, replaces the unobserved regression error terms by some nonparametric estimator $\widehat{\varepsilon}_i$. In a second step, the estimated $\widehat{\varepsilon}_i$'s are used to estimate nonparametrically $f(\cdot)$, as if they were the true ε_i's. If proceeding so can circumvent the

curse of dimensionality, a challenging issue is to evaluate the impact of the estimated residuals on the final estimator of $f(\cdot)$. Hence one of the contributions of our study is to analyze the effect of the estimation of the residuals on the regression errors p.d.f. Kernel estimators. Next, an optimal choice of the bandwidth used to estimate the residuals is given. Finally, we study the asymptotic normality of the feasible Kernel estimator and its rate-optimality.

The rest of this chapter is organized as follows. Section 3.2 presents ours estimators and proposes an asymptotic normality of the (naive) conditional estimator of the density of the regression error. Sections 3.3 and 3.4 group our assumptions and main results. The conclusion of this chapter is given in Section 3.5, while the proofs of our results are gathered in section 3.6 and in an appendix.

3.2 Some nonparametric estimator of the density of the regression error

To illustrate the potential impact of the dimension d of the X_i's, let us first consider a naive conditional estimator of the p.d.f $f(\cdot)$ of the regression error term ε. Let $\varphi(\cdot|x)$ and $f(\cdot|x)$ be respectively the p.d.f. of Y and ε given $X = x$. Since $f(\epsilon|x) = \varphi(m(x)+\epsilon|x)$, using the independence of X and ε gives

$$f(\epsilon) = f(\epsilon|x) = \varphi\left(m(x) + \epsilon|x\right). \tag{3.2.1}$$

Consider some Kernel functions K_0, K_1 and some bandwidths b_0, h_0 and h_1. The expression (3.2.1) of f suggests to use the Kernel nonparametric estimator

$$\widetilde{f}_n(\epsilon|x) = \frac{\frac{1}{nh_0^d h_1}\sum_{i=1}^n K_0\left(\frac{X_i - x}{h_0}\right) K_1\left(\frac{Y_i - \widehat{m}_n(x) - \epsilon}{h_1}\right)}{\frac{1}{nh_0^d}\sum_{i=1}^n K_0\left(\frac{X_i - x}{h_0}\right)},$$

where $\widehat{m}_n(x)$ is the Nadaraya-Watson (1964) estimator of $m(x)$ defined as

$$\widehat{m}_n(x) = \frac{\sum_{j=1}^n Y_j K_0\left(\frac{X_j - x}{b_0}\right)}{\sum_{j=1}^n K_0\left(\frac{X_j - x}{b_0}\right)}. \tag{3.2.2}$$

The first result presented in this chapter is the following proposition.

Proposition 3.1. *Define*

$$\mu_1(x,\epsilon) = \frac{\partial^2 \varphi\left(x, m(x)+\epsilon\right)}{\partial^2 x}\int z K_0(z) z^\top dz, \quad \mu_2(x,\epsilon) = \frac{\partial^2 \varphi\left(x, m(x)+\epsilon\right)}{\partial^2 y}\int v^2 K_1(v)dv,$$

and suppose that h_0 *decrease to* 0 *such that* $nh_0^{2d}/\ln n \to \infty$, $\ln(1/h_0)/\ln(\ln n) \to \infty$ *and*

$$(\mathbf{A_0}): \quad nh_0^d h_1 \to \infty, \quad \left(\frac{nh_0^d}{h_1}\right)\left(b_0^4 + \frac{\ln n}{nb_0^d}\right) = o(1),$$

when $n \to \infty$. Then under Assumptions $(A_1)-(A_{10})$ given in the next section, we have

$$\sqrt{nh_0^d h_1}\left(\widetilde{f}_n(\epsilon|x)-\overline{\widetilde{f}}_n(\epsilon|x)\right)\xrightarrow{d}\mathcal{N}\left(0,\frac{f(\epsilon|x)}{g(x)}\int\int K_0^2(z)K_1^2(v)dzdv\right),$$

where $g(\cdot)$ is the marginal density of X and

$$\overline{\widetilde{f}}_n(\epsilon|x)=f(\epsilon|x)+\frac{h_0^2\mu_1(x,\epsilon)}{2g(x)}+\frac{h_1^2\mu_2(x,\epsilon)}{2g(x)}+o\left(h_0^2+h_1^2\right).$$

This results suggests that an optimal choice of the bandwidths h_0 and h_1 should achieve the minimum of the asymptotic mean square expansion first order terms

$$AMSE\left(\widetilde{f}_n(\epsilon|x)\right)=\left[\frac{h_0^2\mu_1(x,\epsilon)}{2g(x)}+\frac{h_1^2\mu_2(x,\epsilon)}{2g(x)}\right]^2+\frac{f(\epsilon|x)\int K_0^2(z)dz\int K_1^2(v)dv}{nh_0^d h_1 g(x)}.$$

Elementary calculations yield that the resulting optimal bandwidths h_0 and h_1 are all proportional to $n^{-1/(d+5)}$, leading to the exact consistency rate $n^{-2/(d+5)}$ for $\widetilde{f}_n(x|\epsilon)$. In the case $d=1$, this rate is $n^{-1/3}$, which is worst than the rate $n^{-2/5}$ achieved by the optimal Kernel estimator of an univariate density. See Bosq and Lecoutre (1987), Scott (1992), Wand and Jones (1995). Note also that the exponent $2/(d+5)$ decreases to 0 with the dimension d. This indicates a negative impact of the dimension d on the performance of the estimator, the so-called curse of dimensionality. The fact that $\widetilde{f}_n(\epsilon|x)$ is affected by the curse of dimensionality is a consequence of conditioning. Indeed, (3.2.1) identifies the unconditional $f(\epsilon)$ with the conditional distribution of the regression error given the covariate.

To avoid this curse of dimensionality in the nonparametric kernel estimation of $f(\epsilon)$, our approach proposed here builds, in a first step, the estimated residuals

$$\widehat{\varepsilon}_i=Y_i-\widehat{m}_{in},\quad i=1,\ldots,n, \tag{3.2.3}$$

where $\widehat{m}_{in}=\widehat{m}_{in}(X_i)$ is a leave-one out version of the Kernel regression estimator (3.2.2),

$$\widehat{m}_{in}=\frac{\sum_{\substack{j=1\\j\neq i}}^n Y_j K_0\left(\frac{X_j-X_i}{b_0}\right)}{\sum_{\substack{j=1\\j\neq i}}^n K_0\left(\frac{X_j-X_i}{b_0}\right)}. \tag{3.2.4}$$

It is tempting to use, in a second step, the estimated $\widehat{\varepsilon}_i$ as if they were the true residuals ε_i. This would ignore that the $\widehat{m}_n(X_i)$'s can deliver severely biased estimations of the $m(X_i)$'s for those X_i which are close to the boundaries of the support $\mathcal{X}$ of the covariate distribution. To that aim, our proposed estimator trims the observations X_i outside an inner subset $\mathcal{X}_0$ of $\mathcal{X}$,

$$\widehat{f}_{1n}(\epsilon)=\frac{1}{b_1\sum_{i=1}^n \mathbb{1}\left(X_i\in\mathcal{X}_0\right)}\sum_{i=1}^n \mathbb{1}\left(X_i\in\mathcal{X}_0\right)K_1\left(\frac{\widehat{\varepsilon}_i-\epsilon}{b_1}\right). \tag{3.2.5}$$

This estimator is the so-called two-steps Kernel estimator of $f(\epsilon)$. In principle, it would be possible to assume that $\mathcal{X}_0$ grows to $\mathcal{X}$ with a negligible rate compared to the bandwidth b_1. This would

give an estimator close to the more natural Kernel estimator $\sum_{i=1}^{n} K\left((\widehat{\varepsilon}_i - \epsilon)/b_1\right)/(nb_1)$. However, in the rest of the paper, a fixed subset $\mathcal{X}_0$ will be considered for the sake of simplicity.

Observe that the two steps Kernel estimator $\widehat{f}_{1n}(\epsilon)$ is a feasible estimator in the sense that it does not depend on any unknown quantity, as desirable in practice. This contrasts with the unfeasible ideal Kernel estimator

$$\widetilde{f}_{1n}(\epsilon) = \frac{1}{b_1 \sum_{i=1}^{n} \mathbb{1}\left(X_i \in \mathcal{X}_0\right)} \sum_{i=1}^{n} \mathbb{1}\left(X_i \in \mathcal{X}_0\right) K_1\left(\frac{\varepsilon_i - \epsilon}{b_1}\right), \tag{3.2.6}$$

which depends in particular on the unknown regression error terms. It is however intuitively clear that $\widehat{f}_{1n}(\epsilon)$ and $\widetilde{f}_{1n}(\epsilon)$ should be closed, as illustrated by the results of the next section.

3.3 Assumptions

The following assumptions are used in our mains results.

($\mathbf{A_1}$) *The support $\mathcal{X}$ of X is a compact subset of $\mathbb{R}^d$ and $\mathcal{X}_0$ is an inner closed subset of $\mathcal{X}$ with non empty interior,*

($\mathbf{A_2}$) *the p.d.f. $g(\cdot)$ of the i.i.d. covariates X, X_i is strictly positive over $\mathcal{X}_0$, and has continuous second order partial derivatives over $\mathcal{X}$,*

($\mathbf{A_3}$) *the regression function $m(\cdot)$ has continuous second order partial derivatives over $\mathcal{X}$,*

($\mathbf{A_4}$) *the i.i.d. centered error regression terms $\varepsilon, \varepsilon_i$'s, have finite 6th moments, and are independent of the covariates X, X_i's,*

($\mathbf{A_5}$) *the probability density function $f(\cdot)$ has bounded continuous second order derivatives over $\mathbb{R}$ and satisfies, for $h_p(e) = e^p f(e)$, $\sup_{e\in\mathbb{R}} |h_p^{(k)}(e)| < \infty$, $p \in [0,2]$, $k \in [0,2]$,*

($\mathbf{A_6}$) *the p.d.f φ of (X, Y) has bounded continuous second order partial derivatives over $\mathbb{R}^d \times \mathbb{R}$,*

($\mathbf{A_7}$) *the Kernel K_0 is symmetric, continuous over $\mathbb{R}^d$ with support contained in $[-1/2, 1/2]^d$ and $\int K_0(z)dz = 1$,*

($\mathbf{A_8}$) *the Kernel K_1 has a compact support, is three times continuously differentiable over $\mathbb{R}$, and satisfies $\int K_1(v)dv = 1$ and $\int vK_1(v)dv = 0$,*

($\mathbf{A_9}$) *the bandwidth b_0 decreases to 0 and satisfies, for $d^* = \sup\{d+2, 2d\}$, $nb_0^{d^*}/\ln n \to \infty$ and $\ln(1/b_0)/\ln(\ln n) \to \infty$ when $n \to \infty$,*

($\mathbf{A_{10}}$) *the bandwidth b_1 decreases to 0 and satisfies $n^{(d+8)} b_1^{7(d+4)} \to \infty$ when $n \to \infty$.*

Assumptions (A_2), (A_3), (A_5) and (A_6) impose that all the functions to be estimated nonparametrically have two bounded derivatives. Consequently the conditions $\int zK_0(z)dz = 0$ and $\int vK_1(v)dv = 0$, as assumed in (A_7) and (A_8), represent standard conditions ensuring that the

bias of the resulting nonparametric estimators (3.2.2) and (3.2.6) are of order b_0^2 and b_1^2. Assumption (A_4) states independence between the regression error terms and the covariates, which is the main condition for (3.2.1) to hold. The differentiability of K_1 imposed in (A_8) is more specific to our two-steps estimation method. Assumption (A_8) is used to expand the two-steps Kernel estimator $\widehat{f}_{1n}$ in (3.2.5) around the unfeasible one $\widetilde{f}_{1n}$ from (3.2.6), using the residual error estimation $\widehat{\varepsilon}_i - \varepsilon_i$'s and the derivatives of K_1 up to third order. Assumption (A_9) is useful for obtaining the uniform convergence of the regression estimator $\widehat{m}_n$ defined in (3.2.2) (see for instance Einmahl and Mason, 2005), and also gives a similar consistency result for the leave-one-out estimator $\widehat{m}_{in}$ in (3.2.4). Assumption (A_{10}) is needed in the study of the difference between the feasible estimator $\widehat{f}_{1n}$ and the unfeasible estimator $\widetilde{f}_{1n}$.

3.4 Main results

This section is devoted to our main results. The first result we give here concerns the pointwise consistency of the nonparamatric Kernel estimator $\widehat{f}_{1n}$ of the density f. Next, the optimal first-step and second-step bandwidths used to estimated f are proposed. We finish this section by establishing an asymptotic normality for the estimator $\widehat{f}_{1n}$.

3.4.1 Pointwise weak consistency

The next result gives the order of the difference between the feasible estimator and the theoretical density of the regression error at a fixed point ϵ.

Theorem 3.1. *Under $(A_1) - (A_5)$ and $(A_7) - (A_{10})$, we have, when b_0 and b_1 go to 0,*

$$\widehat{f}_{1n}(\epsilon) - f(\epsilon) = O_{\mathbb{P}}\Big(AMSE(b_1) + R_n(b_0, b_1)\Big)^{1/2},$$

where

$$AMSE(b_1) = \mathbb{E}_n\left[\left(\widetilde{f}_{1n}(\epsilon) - f(\epsilon)\right)^2\right] = O_{\mathbb{P}}\left(b_1^4 + \frac{1}{nb_1}\right),$$

and

$$R_n(b_0, b_1) = b_0^4 + \left[\frac{1}{(nb_1^5)^{1/2}} + \left(\frac{b_0^d}{b_1^3}\right)^{1/2}\right]^2\left(b_0^4 + \frac{1}{nb_0^d}\right)^2 + \left[\frac{1}{b_1} + \left(\frac{b_0^d}{b_1^7}\right)^{1/2}\right]^2\left(b_0^4 + \frac{1}{nb_0^d}\right)^3.$$

The result of Theorem 3.1 is based on the evaluation of the difference between $\widehat{f}_{1n}(\epsilon)$ and $\widetilde{f}_{1n}(\epsilon)$. This evaluation gives an indication about the impact of the estimation of the residuals on the nonparametric estimation of the regression error density.

3.4.2 Optimal first-step and second-step bandwidths for the pointwise weak consistency

As shown in the next result, Theorem 3.2 gives some guidelines for the choice of the optimal bandwidth b_0 used in the nonparametric regression errors estimation. As far as we know, the choice of an optimal b_0 has not been addressed before. In what follows, $a_n \asymp b_n$ means that $a_n = O(b_n)$ and $b_n = O(a_n)$, i.e. that there is a constant $C > 0$ such that $|a_n|/C \leq |b_n| \leq C|a_n|$ for n large enough.

Theorem 3.2. *Suppose that* $(A_1) - (A_5)$ *and* $(A_7) - (A_{10})$ *are satisfied, and define*

$$b_0^* = b_0^*(b_1) = \arg\min_{b_0} R_n(b_0, b_1).$$

where the minimization is performed over bandwidth b_0 *fulfilling* (A_9)*. Then the bandwidth* b_0^* *satisfies*

$$b_0^* \asymp \max\left\{\left(\frac{1}{n^2 b_1^3}\right)^{\frac{1}{d+4}}, \left(\frac{1}{n^3 b_1^7}\right)^{\frac{1}{2d+4}}\right\},$$

and we have

$$R_n(b_0^*, b_1) \asymp \max\left\{\left(\frac{1}{n^2 b_1^3}\right)^{\frac{4}{d+4}}, \left(\frac{1}{n^3 b_1^7}\right)^{\frac{4}{2d+4}}\right\}.$$

Our next theorem gives the conditions for which the estimator $\widehat{f}_{1n}(\epsilon)$ reaches the optimal rate $n^{-2/5}$ when b_0 takes the value b_0^*. We prove that for $d \leq 2$, the bandwidth that minimizes the term $AMSE(b_1) + R_n(b_0^*, b_1)$ has the same order as $n^{-1/5}$, yielding the optimal order $n^{-2/5}$ for $(AMSE(b_1) + R_n(b_0^*, b_1))^{1/2}$.

Theorem 3.3. *Assume that* $(A_1) - (A_5)$ *and* $(A_7) - (A_{10})$ *are satisfied, and set*

$$b_1^* = \arg\min_{b_1}\Big(AMSE(b_1) + R_n(b_0^*, b_1)\Big),$$

where $b_0^* = b_0^*(b_1)$ *is defined as in Theorem 3.2. Then*

1. *For* $d \leq 2$*, the bandwidth* b_1^* *satisfies*

$$b_1^* \asymp \left(\frac{1}{n}\right)^{\frac{1}{5}},$$

and we have

$$\Big(AMSE(b_1^*) + R_n(b_0^*, b_1^*)\Big)^{\frac{1}{2}} \asymp \left(\frac{1}{n}\right)^{\frac{2}{5}}.$$

2. *For* $d \geq 3$*,* b_1^* *satisfies*

$$b_1^* \asymp \left(\frac{1}{n}\right)^{\frac{3}{2d+11}},$$

and we have

$$\Big(AMSE(b_1^*) + R_n(b_0^*, b_1^*)\Big)^{\frac{1}{2}} \asymp \left(\frac{1}{n}\right)^{\frac{6}{2d+11}}.$$

The results of Theorem 3.3 show that the rate $n^{-2/5}$ is reachable if and only when $d \leq 2$. These results are derived from Theorem 3.2. This latter indicates that if b_1 is proportional to $n^{-1/5}$, the bandwidth b_0^* has the same order as

$$\max\left\{\left(\frac{1}{n}\right)^{\frac{7}{5(d+4)}}, \left(\frac{1}{n}\right)^{\frac{8}{5(2d+4)}}\right\} = \left(\frac{1}{n}\right)^{\frac{8}{5(2d+4)}}.$$

For $d \leq 2$, this order of b_0^* is smaller than the one of the optimal bandwidth b_{0*} obtained for pointwise or mean square estimation of $m(\cdot)$ using a Kernel estimator. In fact, it has been shown in Nadaraya (1989, Chapter 4) that the optimal bandwidth b_{0*} for estimating $m(\cdot)$ is obtained by minimizing the order of the risk function

$$r_n(b_0) = \mathbb{E}\left[\int \mathbb{1}\left(x \in \mathcal{X}\right)\left(\widehat{m}_n(x) - m(x)\right)^2 \widehat{g}_n^2(x)w(x)dx\right],$$

where $\widehat{g}_n(x)$ is a nonparametric Kernel estimator of $g(x)$, and $w(\cdot)$ is a nonnegative weight function, which is bounded and squared integrable on $\mathcal{X}$. If $g(\cdot)$ and $m(\cdot)$ have continuous second order partial derivatives over their supports, Nadaraya (1989, Chapter 4) shows that $r_n(b_0)$ has the same order as $b_0^4 + \left(1/(nb_0^d)\right)$, leading to the optimal bandwidth $\widehat{b}_0 = n^{-1/(d+4)}$ for the convergence of the estimator $\widehat{m}_n(\cdot)$ of $m(\cdot)$ in the set of the square integrable functions on $\mathcal{X}$.

For d=1, the optimal order of b_0^* is $n^{-(1/5)\times(4/3)}$ which goes to 0 slightly faster than $n^{-1/5}$, the optimal order of the bandwidth $\widehat{b}_0$ for the mean square nonparametric estimation of $m(\cdot)$.

For $d = 2$, the optimal order of b_0^* is $n^{-1/5}$. Again this order goes to 0 faster than the order $n^{-1/6}$ of the optimal bandwidth for the nonparametric estimation of the regression function with two covariates.

However, for $d \geq 3$, we note that the order of b_0^* goes to 0 slowly than $\widehat{b}_0$. Hence our results show that optimal $\widehat{m}_n(\cdot)$ for estimating $f(\cdot)$ should use a very small bandwidth b_0. This suggests that $\widehat{m}_n(\cdot)$ should be less biased and should have a higher variance than the optimal Kernel regression estimator of the estimation setup. Such a finding parallels Wang, Cai, Brown and Levine (2008) who show that a similar result hold when estimating the conditional variance of a heteroscedastic regression error term. However Wang et *al.* (2008) do not give the order of the optimal bandwidth to be used for estimating the regression function in their heteroscedastic setup. These results show that estimators of $m(\cdot)$ with smaller bias should be preferred in our framework, compared to the case where the regression function $m(\cdot)$ is the parameter of interest.

3.4.3 Asymptotic normality

We give now an asymptotic normality of the estimator $\widehat{f}_{1n}(\epsilon)$.

Theorem 3.4. *Assume that*

$$(\mathbf{A_{11}}): \quad nb_0^{d+4} = O(1), \quad nb_0^4 b_1 = o(1), \quad nb_0^d b_1^3 \to \infty,$$

when n goes to ∞. Then under $(A_1)-(A_5)$, $(A_7)-(A_{10})$, we have

$$\sqrt{nb_1}\left(\widehat{f}_{1n}(\epsilon)-\overline{f}_{1n}(\epsilon)\right)\xrightarrow{d}\mathcal{N}\left(0,\frac{f(\epsilon)}{\mathbb{P}\left(X\in\mathcal{X}_0\right)}\int K_1^2(v)dv\right),$$

where

$$\overline{f}_{1n}(\epsilon)=f(\epsilon)+\frac{b_1^2}{2}f^{(2)}(\epsilon)\int v^2K_1(v)dv+o\left(b_1^2\right).$$

The result of this theorem shows that the best choice b_1^* for the bandwidth b_1 should achieve the minimum of the Asymptotic Mean Integrated Square Error

$$\text{AMISE}=\frac{b_1^4}{4}\int\left(f^{(2)}(\epsilon)\right)^2d\epsilon\left(\int v^2K_1(v)dv\right)^2+\frac{1}{nb_1\mathbb{P}\left(X\in\mathcal{X}_0\right)}\int K_1^2(v)dv,$$

leading to the optimal bandwidth

$$b_1^*=\left[\frac{\frac{1}{\mathbb{P}(X\in\mathcal{X}_0)}\int K_1^2(v)dv}{\int(f^{(2)}(\epsilon))^2d\epsilon\left(\int v^2K_1(v)dv\right)^2}\right]^{1/5}n^{-1/5}.$$

We also note that for $d\leq 2$, $b_1=b_1^*$ and $b_0=b_0^*$, Theorems 3.3 and 3.2 give

$$b_1\asymp\left(\frac{1}{n}\right)^{\frac{1}{5}},\quad b_0\asymp\left(\frac{1}{n}\right)^{\frac{8}{5(2d+4)}},$$

which yields that

$$nb_0^{d+4}\asymp\left(\frac{1}{n}\right)^{\frac{12-2d}{5(2d+4)}},\quad nb_0^4b_1\asymp\left(\frac{1}{n}\right)^{\frac{16-8d}{5(2d+4)}},\quad nb_0^db_1^3\asymp\left(\frac{1}{n}\right)^{\frac{4d-8}{5(2d+4)}}.$$

This shows that for $d=1$, the Assumption $(\mathbf{A_{11}})$ is realizable with the optimal bandwidths b_0^* and b_1^*. But with these bandwidths, the last constraint of $(\mathbf{A_{11}})$ is not satisfied for $d=2$, since $nb_0^db_1^3$ is bounded when $n\to\infty$.

3.5 Conclusion

The aim of this chapter was to study the nonparametric Kernel estimation of the probability density function of the regression error using the estimated residuals. The difference between the feasible estimator which uses the estimated residuals and the unfeasible one using the true residuals are studied. An optimal choice of the first-step bandwidth used to estimate the residuals is also proposed. Again, an asymptotic normality of the feasible Kernel estimator and its rate-optimality are established. One of the contributions of this paper is the analysis of the impact of the estimated residuals on the regression errors p.d.f. Kernel estimator.

In our setup, the strategy was to use an approach based on a two-steps procedure which, in a first step, replaces the unobserved residuals terms by some nonparametric estimators $\widehat{\varepsilon}_i$. In a second step, the "pseudo-observations"$\widehat{\varepsilon}_i$ are used to estimate the p.d.f $f(\cdot)$, as if they were the true ε_i's. If proceeding so can remedy the curse of dimensionality, a challenging issue was to measure the impact of the estimated residuals on the final estimator of $f(\cdot)$ in the first nonparametric step, and to find the order of the optimal first-step bandwidth b_0. For this choice of b_0, our results indicates that the optimal bandwidth to be used for estimating the regression function $m(\cdot)$ should be smaller than the optimal bandwidth for the mean square estimation of $m(\cdot)$. That is to say, the best estimator $\widehat{m}_n(\cdot)$ of the regression function $m(\cdot)$ needed for estimating $f(\cdot)$ should have a lower bias and a higher variance than the optimal Kernel regression of the estimation setup. With this appropriate choice of b_0, it has been seen that for $d \leq 2$, the nonparametric estimator $\widehat{f}_{1n}(\epsilon)$ of f can reach the optimal rate $n^{-2/5}$, which corresponds to the exact consistency rate reached for the Kernel density estimator of real-valued variable. Hence our main conclusion is that for $d \leq 2$, the estimator $\widehat{f}_{1n}(\epsilon)$ used for estimating $f(\epsilon)$ is not affected by the curse of dimensionality, since there is no negative effect coming from the estimation of the residuals on the final estimator of $f(\epsilon)$.

3.6 Proofs section

Intermediate Lemmas for Proposition 3.1 and Theorem 3.1

Lemma 3.1. *Define, for $x \in \mathcal{X}_0$,*

$$\widehat{g}_n(x) = \frac{1}{nb_0^d} \sum_{i=1}^n K_0\left(\frac{X_i - x}{b_0}\right), \quad \overline{g}_n(x) = \mathbb{E}\left[\widehat{g}_n(x)\right].$$

Then under $(A_1) - (A_2)$, (A_4), (A_7) and (A_9), we have, when b_0 goes to 0,

$$\sup_{x \in \mathcal{X}_0} |\overline{g}_n(x) - g(x)| = O\left(b_0^2\right), \quad \sup_{x \in \mathcal{X}_0} |\widehat{g}_n(x) - \overline{g}_n(x)| = O_{\mathbb{P}}\left(b_0^4 + \frac{\ln n}{nb_0^d}\right)^{1/2},$$

and

$$\sup_{x \in \mathcal{X}_0} \left|\frac{1}{\widehat{g}_n(x)} - \frac{1}{g(x)}\right| = O_{\mathbb{P}}\left(b_0^4 + \frac{\ln n}{nb_0^d}\right)^{1/2}.$$

Lemma 3.2. *Under $(A_1) - (A_4)$, (A_7) and (A_9), we have*

$$\sup_{x \in \mathcal{X}_0} |\widehat{m}_n(x) - m(x)| = O_{\mathbb{P}}\left(b_0^4 + \frac{\ln n}{nb_0^d}\right)^{1/2}.$$

Lemma 3.3. *Define for $(x, y) \in \mathbb{R}^d \times \mathbb{R}$,*

$$f_n(\epsilon|x) = \frac{\frac{1}{nh_0^d h_1} \sum_{i=1}^n K_0\left(\frac{X_i - x}{h_0}\right) K_1\left(\frac{Y_i - m(x) - \epsilon}{h_1}\right)}{\frac{1}{nh_0^d} \sum_{i=1}^n K_0\left(\frac{X_i - x}{h_0}\right)},$$

Then under $(A_1) - (A_3)$, $(A_6) - (A_9)$, we have, when n goes to infinity,

$$\widetilde{f}_n(\epsilon|x) - f_n(\epsilon|x) = o_{\mathbb{P}}\left(\frac{1}{nh_0^d h_1}\right)^{1/2}.$$

Lemma 3.4. *Set, for $(x, y) \in \mathbb{R}^d \times \mathbb{R}$,*

$$\widetilde{\varphi}_{in}(x, y) = \frac{1}{h_0^d h_1} K_0\left(\frac{X_i - x}{h_0}\right) K_1\left(\frac{Y_i - y}{h_1}\right).$$

Then, under $(A_6) - (A_8)$, we have, for x in $\mathcal{X}_0$ and y in $\mathbb{R}$, h_0 and h_1 going to 0, and for some constant $C > 0$,

$$\begin{aligned}
\mathbb{E}\left[\widetilde{\varphi}_{in}(x, y)\right] - \varphi(x, y) &= \frac{h_0^2}{2} \frac{\partial^2 \varphi(x, y)}{\partial^2 x} \int z K_0(z) z^\top dz + \frac{h_1^2}{2} \frac{\partial^2 \varphi(x, y)}{\partial^2 y} \int v^2 K_1(v) dv \\
&\quad + o\left(h_0^2 + h_1^2\right), \\
\operatorname{Var}\left[\widetilde{\varphi}_{in}(x, y)\right] &= \frac{\varphi(x, y)}{h_0^d h_1} \int\int K_0^2(z) K_1^2(v) dv dz + o\left(\frac{1}{h_0^d h_1}\right), \\
\mathbb{E}\left[\left|\widetilde{\varphi}_{in}(x, y) - \mathbb{E}\widetilde{\varphi}_{in}(x, y)\right|^3\right] &\leq \frac{C\varphi(x, y)}{h_0^{2d} h_1^2} \int\int |K_0(z) K_1(v)|^3 dz dv + o\left(\frac{1}{h_0^{2d} h_1^2}\right).
\end{aligned}$$

Lemma 3.5. *Set*

$$f_{in}(\epsilon)=\frac{\mathbb{1}\left(X_i\in\mathcal{X}_0\right)}{b_1\mathbb{P}\left(X\in\mathcal{X}_0\right)}K_1\left(\frac{\varepsilon_i-\epsilon}{b_1}\right).$$

Then under (A_4), (A_5) and (A_8), we have, for b_1 going to 0, and for some constant $C>0$,

$$\begin{aligned}
\mathbb{E}f_{in}(\epsilon) &= f(\epsilon)+\frac{b_1^2}{2}f^{(2)}(\epsilon)\int v^2K_1(v)dv+o\left(b_1^2\right),\\
\mathrm{Var}\left(f_{in}(\epsilon)\right) &= \frac{f(\epsilon)}{b_1\mathbb{P}\left(X\in\mathcal{X}_0\right)}\int K_1^2(v)dv+o\left(\frac{1}{b_1}\right),\\
\mathbb{E}\left|f_{in}(\epsilon)-\mathbb{E}f_{in}(\epsilon)\right|^3 &\leq \frac{Cf(\epsilon)}{b_1^2\mathbb{P}^2\left(X\in\mathcal{X}_0\right)}\int\left|K_1(v)\right|^3dv+o\left(\frac{1}{b_1^2}\right).
\end{aligned}$$

Lemma 3.6. *Define*

$$\begin{aligned}
S_n &= \sum_{i=1}^n \mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left(\widehat{m}_{in}-m(X_i)\right)K_1^{(1)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right),\\
T_n &= \sum_{i=1}^n \mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left(\widehat{m}_{in}-m(X_i)\right)^2K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right),\\
R_n &= \sum_{i=1}^n \mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left(\widehat{m}_{in}-m(X_i)\right)^3\int_0^1(1-t)^2K_1^{(3)}\left(\frac{\varepsilon_i-t(\widehat{m}_{in}-m(X_i))-\epsilon}{b_1}\right)dt.
\end{aligned}$$

Then under $(A_1)-(A_5)$ and $(A_7)-(A_{10})$, we have, for b_0 and b_1 small enough,

$$\begin{aligned}
S_n &= O_{\mathbb{P}}\left[b_0^2\left(nb_1^2+(nb_1)^{1/2}\right)+\left(nb_1^4+\frac{b_1}{b_0^d}\right)^{1/2}\right],\\
T_n &= O_{\mathbb{P}}\left[\left(nb_1^3+(nb_1)^{1/2}+\left(n^2b_0^db_1^3\right)^{1/2}\right)\left(b_0^4+\frac{1}{nb_0^d}\right)\right],\\
R_n &= O_{\mathbb{P}}\left[\left(nb_1^3+\left(n^2b_0^db_1\right)^{1/2}\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^{3/2}\right].
\end{aligned}$$

Lemma 3.7. *Under (A_5) and (A_8) we have, for some constant $C>0$, and for any ϵ in $\mathbb{R}$ and $p\in[0,2]$,*

$$\left|\int K_1^{(1)}\left(\frac{e-\epsilon}{b_1}\right)^2e^pf(e)de\right|\leq Cb_1,\qquad \left|\int K_1^{(1)}\left(\frac{e-\epsilon}{b_1}\right)e^pf(e)de\right|\leq Cb_1^2,\tag{3.6.1}$$

$$\left|\int K_1^{(2)}\left(\frac{e-\epsilon}{b_1}\right)^2e^pf(e)de\right|\leq Cb_1,\qquad \left|\int K_1^{(2)}\left(\frac{e-\epsilon}{b_1}\right)e^pf(e)de\right|\leq Cb_1^3,\tag{3.6.2}$$

$$\left|\int K_1^{(3)}\left(\frac{e-\epsilon}{b_1}\right)^2e^pf(e)de\right|\leq Cb_1,\qquad \left|\int K_1^{(3)}\left(\frac{e-\epsilon}{b_1}\right)e^pf(e)de\right|\leq Cb_1^3.\tag{3.6.3}$$

Lemma 3.8. *Set*

$$\beta_{in}=\frac{\mathbb{1}\left(X_i\in\mathcal{X}_0\right)}{nb_0^d\widehat{g}_{in}}\sum_{j=1,j\neq i}^n\left(m(X_j)-m(X_i)\right)K_0\left(\frac{X_j-X_i}{b_0}\right).$$

Then, under $(A_1)-(A_5)$ and $(A_7)-(A_{10})$, we have, when b_0 and b_1 go to 0,

$$\sum_{i=1}^{n}\beta_{in}K_1^{(1)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)=O_{\mathbb{P}}\left(b_0^2\right)\left(nb_1^2+(nb_1)^{1/2}\right).$$

Lemma 3.9. *Set*

$$\Sigma_{in}=\frac{\mathbb{1}\left(X_i\in\mathcal{X}_0\right)}{nb_0^d\widehat{g}_{in}}\sum_{j=1,j\neq i}^{n}\varepsilon_jK_0\left(\frac{X_j-X_i}{b_0}\right).$$

Then, under $(A_1)-(A_5)$ and $(A_7)-(A_{10})$, we have

$$\sum_{i=1}^{n}\Sigma_{in}K_1^{(1)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)=O_{\mathbb{P}}\left(nb_1^4+\frac{b_1}{b_0^d}\right)^{1/2}.$$

Lemma 3.10. *Let $\mathbb{E}_n[\cdot]$ be the conditional mean given $X_1,\ldots,X_n$. Then under $(A_1)-(A_5)$ and $(A_7)-(A_9)$, we have, for b_0 going to 0,*

$$\begin{aligned}\sup_{1\leq i\leq n}\mathbb{E}_n\left[\mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left(\widehat{m}_{in}-m(X_i)\right)^4\right] &= O_{\mathbb{P}}\left(b_0^4+\frac{1}{nb_0^d}\right)^2,\\ \sup_{1\leq i\leq n}\mathbb{E}_n\left[\mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left(\widehat{m}_{in}-m(X_i)\right)^6\right] &= O_{\mathbb{P}}\left(b_0^4+\frac{1}{nb_0^d}\right)^3.\end{aligned}$$

Lemma 3.11. *Assume that (A_4) and (A_7) hold. Then, for any $1\leq i\neq j\leq n$, and for any ϵ in $\mathbb{R}$,*

$$\left(\widehat{m}_{in}-m(X_i),\varepsilon_i\right)\text{ and }\left(\widehat{m}_{jn}-m(X_j),\varepsilon_j\right)$$

are independent given $X_1,\ldots,X_n$, provided that $\|X_i-X_j\|\geq Cb_0$, for some constant $C>0$.

Lemma 3.12. *Let $\mathrm{Var}_n(\cdot)$ and $Cov_n(\cdot)$ be respectively the conditional variance and the conditional covariance given $X_1,\ldots,X_n$, and set*

$$\zeta_{in}=\mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left(\widehat{m}_{in}-m(X_i)\right)^2K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right).$$

Then under $(A_1)-(A_5)$ and $(A_7)-(A_9)$, we have, for n going to infinity,

$$\begin{aligned}\sum_{i=1}^{n}\mathrm{Var}_n\left(\zeta_{in}\right) &= O_{\mathbb{P}}\left(nb_1\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^2,\\ \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathrm{Cov}_n\left(\zeta_{in},\zeta_{jn}\right) &= O_{\mathbb{P}}\left(n^2b_0^db_1^{7/2}\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^2.\end{aligned}$$

All these lemmas are proved in Appendix A.

Proof of Proposition 3.1

Define $f_n(\epsilon|x)$ as in Lemma 3.3, and note that by this lemma, we have

$$\widetilde{f}_n(\epsilon|x) = f_n(\epsilon|x) + o_{\mathbb{P}}\left(\frac{1}{nh_0^d h_1}\right)^{1/2}. \tag{3.6.4}$$

The asymptotic distribution of the first term in (3.6.4) is derived by applying the Lyapounov Central Limit Theorem for triangular arrays (see e.g Billingsley 1968, Theorem 7.3). Define for $x \in \mathcal{X}_0$ and $y \in \mathbb{R}$,

$$\widetilde{\varphi}_n(x,y) = \frac{1}{nh_0^d h_1}\sum_{i=1}^n K_0\left(\frac{X_i - x}{h_0}\right) K_1\left(\frac{Y_i - y}{h_1}\right), \quad \widetilde{g}_n(x) = \frac{1}{nh_0^d}\sum_{i=1}^n K_0\left(\frac{X_i - x}{h_0}\right),$$

and observe that

$$f_n(\epsilon|x) = \frac{\widetilde{\varphi}_n\left(x, m(x)+\epsilon\right)}{\widetilde{g}_n(x)}. \tag{3.6.5}$$

Let now $\widetilde{\varphi}_{in}(x,y)$ be as in Lemma 3.4, and note that

$$\widetilde{\varphi}_n(x,y) = \frac{1}{n}\sum_{i=1}^n \Big(\widetilde{\varphi}_{in}(x,y) - \mathbb{E}\left[\widetilde{\varphi}_{in}(x,y)\right]\Big) + \mathbb{E}\left[\widetilde{\varphi}_{1n}(x,y)\right]. \tag{3.6.6}$$

The second and third inequalities in Lemma 3.4 give, since $h_0^d h_1$ goes to 0,

$$\frac{\sum_{i=1}^n \mathbb{E}\left|\widetilde{\varphi}_{in}(x,y) - \mathbb{E}\widetilde{\varphi}_{in}(x,y)\right|^3}{\left(\sum_{i=1}^n \operatorname{Var}\left[\widetilde{\varphi}_{in}(x,y)\right]\right)^3} \leq \frac{\frac{Cn\varphi(x,y)}{h_0^{2d}h_1^2}\int\int \left|K_0(z)K_1(v)\right|^3 dzdv + o\left(\frac{n}{h_0^{2d}h_1^2}\right)}{\left(\frac{n\varphi(x,y)}{h_0^d h_1}\int\int K_0^2(z)K_1^2(v)dvdz + o\left(\frac{n}{h_0^d h_1}\right)\right)^3} = O(h_0^d h_1) = o(1).$$

Hence the Lyapounov Central Limit Theorem gives, since $nh_0^d h_1$ diverges under $(\mathbf{A_0})$,

$$\frac{\sum_{i=1}^n \left\{\widetilde{\varphi}_{in}(x,y) - \mathbb{E}\left[\widetilde{\varphi}_{in}(x,y)\right]\right\}}{\left(\sum_{i=1}^n \operatorname{Var}\left[\widetilde{\varphi}_{in}(x,y)\right]\right)^{1/2}} \xrightarrow{d} \mathcal{N}(0,1),$$

so that

$$\frac{\sqrt{nh_0^d h_1}}{n}\sum_{i=1}^n \Big(\widetilde{\varphi}_{in}(x,y) - \mathbb{E}\left[\widetilde{\varphi}_{in}(x,y)\right]\Big) \xrightarrow{d} \mathcal{N}\left(0, \varphi(x,y)\int\int K_0^2(z)K_1^2(v)dzdv\right). \tag{3.6.7}$$

Further, a similar proof as the one of Lemma 3.1 gives

$$\frac{1}{\widetilde{g}_n(x)} = \frac{1}{g(x)} + O_{\mathbb{P}}\left(h_0^4 + \frac{\ln n}{nh_0^d}\right)^{1/2}. \tag{3.6.8}$$

Hence by this equality, it follows that, taking $y = m(x) + \epsilon$ in (3.6.7), and by (3.6.4)-(3.6.6),

$$\sqrt{nh_0^d h_1}\left(\widetilde{f}_n(\epsilon|x) - \overline{f}_n(\epsilon|x)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{f(\epsilon|x)}{g(x)}\int\int K_0^2(z)K_1^2(v)dzdv\right),$$

where

$$\overline{f}_n(\epsilon|x) = \frac{\mathbb{E}\left[\widetilde{\varphi}_{1n}\left(x, m(x)+\epsilon\right)\right]}{\widetilde{g}_n(x)}.$$

This yields the result of Proposition 3.1, since the first equality of Lemma 3.4 and (3.6.8) yield, for h_0 and h_1 small enough,

$$\begin{aligned}\overline{f}_n(\epsilon|x) &= f(\epsilon|x) + \frac{h_0^2}{2g(x)}\frac{\partial^2\varphi\left(x, m(x)+\epsilon\right)}{\partial^2 x}\int zK_0(z)z^\top dz \\ &+ \frac{h_1^2}{2g(x)}\frac{\partial^2\varphi(x, m(x)+\epsilon)}{\partial^2 y}\int v^2K_1(v)dv + o\left(h_0^2 + h_1^2\right).\square\end{aligned}$$

Proof of Theorem 3.1

The proof of the theorem is based upon the following equalities :

$$\begin{aligned}\widehat{f}_{1n}(\epsilon)-\widetilde{f}_{1n}(\epsilon) &= O_{\mathbb{P}}\left[b_0^2+\left(\frac{1}{n}+\frac{1}{n^2b_0^db_1^3}\right)^{1/2}\right]+O_{\mathbb{P}}\left[\frac{1}{(nb_1^5)^{1/2}}+\left(\frac{b_0^d}{b_1^3}\right)^{1/2}\right]\left(b_0^4+\frac{1}{nb_0^d}\right)\\ &\quad+O_{\mathbb{P}}\left[\frac{1}{b_1}+\left(\frac{b_0^d}{b_1^7}\right)^{1/2}\right]\left(b_0^4+\frac{1}{nb_0^d}\right)^{3/2}, \qquad (3.6.9)\end{aligned}$$

and

$$\widetilde{f}_{1n}(\epsilon)-f(\epsilon)=O_{\mathbb{P}}\left(b_1^4+\frac{1}{nb_1}\right)^{1/2}. \qquad (3.6.10)$$

Indeed, since $\widehat{f}_{1n}(\epsilon)-f(\epsilon)=\left(\widetilde{f}_{1n}(\epsilon)-f(\epsilon)\right)+\widehat{f}_{1n}(\epsilon)-\widetilde{f}_{1n}(\epsilon)$, it then follows by (3.6.10) and (3.6.9) that

$$\begin{aligned}\widehat{f}_{1n}(\epsilon)-f(\epsilon) &= O_{\mathbb{P}}\left[b_1^4+\frac{1}{nb_1}+b_0^4+\frac{1}{n}+\frac{1}{n^2b_0^db_1^3}+\left(\frac{1}{(nb_1^5)^{1/2}}+\left(\frac{b_0^d}{b_1^3}\right)^{1/2}\right)^2\left(b_0^4+\frac{1}{nb_0^d}\right)^2\right]^{1/2}\\ &\quad+O_{\mathbb{P}}\left[\left(\frac{1}{b_1}+\left(\frac{b_0^d}{b_1^7}\right)^{1/2}\right)^2\left(b_0^4+\frac{1}{nb_0^d}\right)^3\right]^{1/2}.\end{aligned}$$

This yields the result of the Theorem, since under (A_9) and (A_{10}), we have

$$\frac{1}{n}=O\left(\frac{1}{nb_1}\right),\quad \frac{1}{n^2b_0^db_1^3}=O\left(\frac{b_0^d}{b_1^3}\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^2.$$

Hence, it remains to prove (3.6.9) and (3.6.10). For this, define S_n, R_n and T_n as in Lemma 3.6. Since $\widehat{\varepsilon}_i-\varepsilon_i=-\left(\widehat{m}_{in}-m(X_i)\right)$ and that K_1 is three times continuously differentiable under (A_8), the third-order Taylor expansion with integral remainder gives

$$\begin{aligned}\widehat{f}_{1n}(\epsilon)-\widetilde{f}_{1n}(\epsilon) &= \frac{1}{b_1\sum_{i=1}^n\mathbb{1}(X_i\in\mathcal{X}_0)}\sum_{i=1}^n\mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left[K_1\left(\frac{\widehat{\varepsilon}_i-\epsilon}{b_1}\right)-K_1\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\\ &= -\frac{1}{b_1\sum_{i=1}^n\mathbb{1}(X_i\in\mathcal{X}_0)}\left(\frac{S_n}{b_1}-\frac{T_n}{2b_1^2}+\frac{R_n}{2b_1^3}\right).\end{aligned}$$

Therefore, since

$$\sum_{i=1}^n\mathbb{1}\left(X_i\in\mathcal{X}_0\right)=n\left(\mathbb{P}\left(X\in\mathcal{X}_0\right)+o_{\mathbb{P}}(1)\right),$$

by the Law of large numbers, Lemma 3.6 then gives

$$\begin{aligned}&\widehat{f}_{1n}(\epsilon)-\widetilde{f}_{1n}(\epsilon)=O_{\mathbb{P}}\left(\frac{1}{nb_1^2}\right)S_n+O_{\mathbb{P}}\left(\frac{1}{nb_1^3}\right)T_n+O_{\mathbb{P}}\left(\frac{1}{nb_1^4}\right)R_n\\ &= O_{\mathbb{P}}\left[b_0^2\left(1+\frac{1}{(nb_1^3)^{1/2}}\right)+\left(\frac{1}{n}+\frac{1}{n^2b_0^db_1^3}\right)^{1/2}\right]\\ &\quad+O_{\mathbb{P}}\left[1+\frac{1}{(nb_1^5)^{1/2}}+\left(\frac{b_0^d}{b_1^3}\right)^{1/2}\right]\left(b_0^4+\frac{1}{nb_0^d}\right)+O_{\mathbb{P}}\left[\frac{1}{b_1}+\left(\frac{b_0^d}{b_1^7}\right)^{1/2}\right]\left(b_0^4+\frac{1}{nb_0^d}\right)^{3/2}.\end{aligned}$$

This yields (3.6.9), since under (A_9) and (A_{10}), we have $b_0 \to 0$, $nb_0^{d+2} \to \infty$ and $nb_1^3 \to \infty$, so that

$$\begin{aligned} b_0^2 \left(1 + \frac{1}{(nb_1^3)^{1/2}}\right) &\asymp O\left(b_0^2\right), \quad \left(b_0^4 + \frac{1}{nb_0^d}\right) = O\left(b_0^2\right), \\ \left[1 + \frac{1}{(nb_1^5)^{1/2}} + \left(\frac{b_0^d}{b_1^3}\right)^{1/2}\right]\left(b_0^4 + \frac{1}{nb_0^d}\right) &= O\left(b_0^2\right) + \left[\frac{1}{(nb_1^5)^{1/2}} + \left(\frac{b_0^d}{b_1^3}\right)^{1/2}\right]\left(b_0^4 + \frac{1}{nb_0^d}\right). \end{aligned}$$

For (3.6.10), note that

$$\mathbb{E}_n\left[\left(\widetilde{f}_{1n}(\epsilon) - f(\epsilon)\right)^2\right] = \mathrm{Var}_n\left(\widetilde{f}_{1n}(\epsilon)\right) + \left(\mathbb{E}_n\left[\widetilde{f}_{1n}(\epsilon)\right] - f(\epsilon)\right)^2, \tag{3.6.11}$$

with, using (A_4),

$$\mathrm{Var}_n\left(\widetilde{f}_{1n}(\epsilon)\right) = \frac{1}{\left(b_1 \sum_{i=1}^n \mathbb{1}\left(X_i \in \mathcal{X}_0\right)\right)^2} \sum_{i=1}^n \mathbb{1}\left(X_i \in \mathcal{X}_0\right) \mathrm{Var}\left[K_1\left(\frac{\varepsilon - \epsilon}{b_1}\right)\right].$$

Therefore, since the Cauchy-Schwarz inequality gives

$$\mathrm{Var}\left[K_1\left(\frac{\varepsilon - \epsilon}{b_1}\right)\right] \le \mathbb{E}\left[K_1^2\left(\frac{\varepsilon - \epsilon}{b_1}\right)\right] \le b_1 \int K_1^2(v) f(\epsilon + b_1 v) dv,$$

this bound and the equality above yield, under (A_5) and (A_8),

$$\mathrm{Var}_n\left(\widetilde{f}_{1n}(\epsilon)\right) \le \frac{C}{b_1 \sum_{i=1}^n \mathbb{1}\left(X_i \in \mathcal{X}_0\right)} = O_{\mathbb{P}}\left(\frac{1}{nb_1}\right). \tag{3.6.12}$$

For the second term in (3.6.11), we have

$$\mathbb{E}_n\left[\widetilde{f}_{1n}(\epsilon)\right] = \frac{1}{b_1 \sum_{i=1}^n \mathbb{1}\left(X_i \in \mathcal{X}_0\right)} \sum_{i=1}^n \mathbb{1}\left(X_i \in \mathcal{X}_0\right) \mathbb{E}\left[K_1\left(\frac{\varepsilon - \epsilon}{b_1}\right)\right]. \tag{3.6.13}$$

By (A_8), K_1 is symmetric, has a compact support, with $\int vK_1(v) = 0$ and $\int K_1(v)dv = 1$. Therefore, since under (A_5) f has bounded continuous second order derivatives, this yields for some $\theta = \theta(\epsilon, b_1 v)$,

$$\begin{aligned} \mathbb{E}\left[K_1\left(\frac{\varepsilon - \epsilon}{b_1}\right)\right] &= b_1 \int K_1(v) f(\epsilon + b_1 v) dv \\ &= b_1 \int K_1(v)\left[f(\epsilon) + b_1 v f^{(1)}(\epsilon) + \frac{b_1^2 v^2}{2} f^{(2)}(\epsilon + \theta b_1 v)\right] dv \\ &= b_1 f(\epsilon) + \frac{b_1^3}{2} \int v^2 K_1(v) f^{(2)}(\epsilon + \theta b_1 v) dv. \end{aligned}$$

Hence this equality and (3.6.13) give

$$\mathbb{E}_n\left[\widetilde{f}_{1n}(\epsilon)\right] = f(\epsilon) + \frac{b_1^2}{2} \int v^2 K_1(v) f^{(2)}(\epsilon + \theta b_1 v) dv,$$

so that

$$\left(\mathbb{E}_n\left[\widetilde{f}_{1n}(\epsilon)\right] - f(\epsilon)\right)^2 = O_{\mathbb{P}}\left(b_1^4\right).$$

Combining this result with (3.6.12) and (3.6.11), we obtain, by the Tchebychev inequality,

$$\widetilde{f}_{1n}(\epsilon) - f(\epsilon) = O_{\mathbb{P}}\left(b_1^4 + \frac{1}{nb_1}\right)^{1/2}.$$

This proves (3.6.10), and then achieves the proof of the theorem. □

Proof of Theorem 3.2

Recall that

$$R_n(b_0,b_1) = b_0^4 + \left[\frac{1}{(nb_1^5)^{1/2}} + \left(\frac{b_0^d}{b_1^3}\right)^{1/2}\right]^2 \left(b_0^4 + \frac{1}{nb_0^d}\right)^2 + \left[\frac{1}{b_1} + \left(\frac{b_0^d}{b_1^7}\right)^{1/2}\right]^2 \left(b_0^4 + \frac{1}{nb_0^d}\right)^3,$$

and note that

$$\left(\frac{1}{n^2 b_1^3}\right)^{\frac{1}{d+4}} = \max\left\{\left(\frac{1}{n^2 b_1^3}\right)^{\frac{1}{d+4}}, \left(\frac{1}{n^3 b_1^7}\right)^{\frac{1}{2d+4}}\right\}$$

if and only if $n^{4-d} b_1^{d+16} \to \infty$. To find the order of b_0^*, we shall deal with the cases $nb_0^{d+4} \to \infty$ and $nb_0^{d+4} = O(1)$.

First assume that $nb_0^{d+4} \to \infty$. More precisely, we suppose that b_0 is in $\left[(u_n/n)^{1/(d+4)}, +\infty\right)$, where $u_n \to \infty$. Since $1/(nb_0^d) = O(b_0^4)$ for all these b_0, we have

$$\left(b_0^4 + \frac{1}{nb_0^d}\right)^2 \asymp \left(b_0^4\right)^2, \quad \left(b_0^4 + \frac{1}{nb_0^d}\right)^3 \asymp \left(b_0^4\right)^3.$$

Hence the order of b_0^* is computed by minimizing the function

$$b_0 \to b_0^4 + \left[\frac{1}{(nb_1^5)^{1/2}} + \left(\frac{b_0^d}{b_1^3}\right)^{1/2}\right]^2 \left(b_0^4\right)^2 + \left[\frac{1}{b_1} + \left(\frac{b_0^d}{b_1^7}\right)^{1/2}\right]^2 \left(b_0^4\right)^3.$$

Since this function is increasing with b_0, the minimum of $R_n(\cdot, b_1)$ is achieved for $b_{0*} = (u_n/n)^{1/(d+4)}$. We shall prove later on that this choice of b_{0*} is irrelevant compared to the one arising when $nb_0^{d+4} = O(1)$.

Consider now the case $nb_0^{d+4} = O(1)$ i.e $b_0^4 = O\left(1/(nb_0^d)\right)$. This gives

$$\left[\frac{1}{(nb_1^5)^{1/2}} + \left(\frac{b_0^d}{b_1^3}\right)^{1/2}\right]^2 \left(b_0^4 + \frac{1}{nb_0^d}\right)^2 \asymp \left(\frac{1}{nb_1^5} + \frac{b_0^d}{b_1^3}\right)\left(\frac{1}{n^2 b_0^{2d}}\right),$$

$$\left[\frac{1}{b_1} + \left(\frac{b_0^d}{b_1^7}\right)^{1/2}\right]^2 \left(b_0^4 + \frac{1}{nb_0^d}\right)^3 \asymp \left(\frac{1}{b_1^2} + \frac{b_0^d}{b_1^7}\right)\left(\frac{1}{n^3 b_0^{3d}}\right).$$

Moreover if $nb_0^d b_1^4 \to \infty$, we have, since $nb_0^{2d} \to \infty$ under (A_9),

$$\left(\frac{1}{nb_1^5} + \frac{b_0^d}{b_1^3}\right)\left(\frac{1}{n^2 b_0^{2d}}\right) \asymp \frac{b_0^d}{b_1^3}\left(\frac{1}{n^2 b_0^{2d}}\right), \quad \left(\frac{1}{b_1^2} + \frac{b_0^d}{b_1^7}\right)\left(\frac{1}{n^3 b_0^{3d}}\right) = O\left(\frac{b_0^d}{b_1^3}\right)\left(\frac{1}{n^2 b_0^{2d}}\right).$$

Hence the order of b_0^* is obtained by finding the minimum of the function $b_0^4 + \left(1/n^2 b_0^d b_1^3\right)$. The minimization of this function gives a solution b_0 such that

$$b_0 \asymp \left(\frac{1}{n^2 b_1^3}\right)^{\frac{1}{d+4}}, \quad R_n(b_0,b_1) \asymp \left(\frac{1}{n^2 b_1^3}\right)^{\frac{4}{d+4}}.$$

This value satisfies the constraints $nb_0^{d+4} = O(1)$ and $nb_0^d b_1^4 \to \infty$ when $n^{4-d} b_1^{d+16} \to \infty$.

If now $nb_0^{d+4} = O(1)$ but $nb_0^d b_1^4 = O(1)$, we have, since $nb_0^{2d} \to \infty$,

$$\frac{1}{nb_1^5}\left(\frac{1}{n^2 b_0^{2d}}\right) = O\left(\frac{b_0^d}{b_1^7}\right)\left(\frac{1}{n^3 b_0^{3d}}\right), \quad \frac{1}{b_1^2}\left(\frac{1}{n^3 b_0^{3d}}\right) = O\left(\frac{b_0^d}{b_1^3}\right)\left(\frac{1}{n^2 b_0^{2d}}\right) = O\left(\frac{b_0^d}{b_1^7}\right)\left(\frac{1}{n^3 b_0^{3d}}\right).$$

In this case, b_0^* is obtained by minimizing the function $b_0^4 + \left(1/n^3 b_0^{2d} b_1^7\right)$, for which the solution b_0 verifies

$$b_0 \asymp \left(\frac{1}{n^3 b_1^7}\right)^{\frac{1}{2d+4}}, \quad R_n(b_0, b_1) \asymp \left(\frac{1}{n^3 b_1^7}\right)^{\frac{4}{2d+4}}.$$

This solution fulfills the constraint $n b_0^d b_1^4 = O(1)$ when $n^{4-d} b_1^{d+16} = O(1)$. Hence we can conclude that for $b_0^4 = O\left(1/(n b_0^d)\right)$, the bandwidth b_0^* satisfies

$$b_0^* \asymp \max\left\{\left(\frac{1}{n^2 b_1^3}\right)^{\frac{1}{d+4}}, \left(\frac{1}{n^3 b_1^7}\right)^{\frac{1}{2d+4}}\right\},$$

which leads to

$$R_n\left(b_0^*, b_1\right) \asymp \max\left\{\left(\frac{1}{n^2 b_1^3}\right)^{\frac{4}{d+4}}, \left(\frac{1}{n^3 b_1^7}\right)^{\frac{4}{2d+4}}\right\}.$$

We need now to compare the solution b_0^* to the candidate $b_{0*} = (u_n/n)^{1/(d+4)}$ obtained when $n b_0^{d+4} \to \infty$. For this, we must do a comparison between the orders of $R_n(b_0^*, b_1)$ and $R_n(b_{0*}, b_1)$. Since $R_n(b_0, b_1) \geq b_0^4$, we have $R_n(b_{0*}, b_1) \geq (u_n/n)^{4/(d+4)}$, so that, for n large enough,

$$\begin{aligned} \frac{R_n(b_0^*, b_1)}{R_n(b_{0*}, b_1)} &\leq C\left[\left(\frac{1}{n^2 b_1^3}\right)^{\frac{1}{d+4}} + \left(\frac{1}{n^3 b_1^7}\right)^{\frac{4}{2d+4}}\right]\left(\frac{n}{u_n}\right)^{\frac{4}{d+4}} \\ &= o(1) + O\left(\frac{1}{u_n}\right)^{\frac{4}{d+4}}\left(\frac{1}{n b_1^{\frac{7(d+4)}{d+8}}}\right)^{\frac{4(d+8)}{(2d+4)(d+4)}} = o(1), \end{aligned}$$

using $u_n \to \infty$ and that $n^{(d+8)} b_1^{7(d+4)} \to \infty$ by (A_{10}). This shows that $R_n(b_0^*, b_1) \leq R_n(b_{0*}, b_1)$ for n large enough. Hence the Theorem is proved, since b_0^* is the best candidate for the minimization of $R_n(\cdot, b_1)$. □

Proof of Theorem 3.3

Recall that Theorem 3.2 gives

$$AMSE(b_1) + R_n(b_0^*, b_1) \asymp r_1(b_1) + r_2(b_1) + r_3(b_1) = F(b_1),$$

where

$$\begin{aligned} r_1(h) &= h^4 + \frac{1}{nh}, \quad \arg\min r_1(h) \asymp n^{-1/5} = h_1^*, \quad \min r_1(h) \asymp (h_1^*)^4 = n^{-4/5}, \\ r_2(h) &= h^4 + \frac{1}{n^{\frac{8}{d+4}} h^{\frac{12}{d+4}}}, \quad \arg\min r_2(h) \asymp n^{-\frac{2}{d+7}} = h_2^*, \quad \min r_3(h) \asymp (h_2^*)^4 = n^{-\frac{8}{d+7}}, \\ r_3(h) &= h^4 + \frac{1}{n^{\frac{12}{2d+4}} h^{\frac{28}{2d+4}}}, \quad \arg\min r_3(h) \asymp n^{-\frac{3}{2d+11}} = h_3^*, \quad \min r_3(h) \asymp (h_3^*)^4 = n^{-\frac{12}{2d+11}}. \end{aligned}$$

Each $r_j(h)$ decreases on $[0, \arg\min r_j(h)]$ and increases on $(\arg\min r_j(h), \infty)$ and that $r_j(h) \asymp h^4$ on $(\arg\min r_j(h), \infty)$. Moreover $\min r_2(h) = o\left(r_3(h)\right)$ and $h_2^* = o\left(h_3^*\right)$ for all possible dimension d, so that $\min\{r_2(h) + r_3(h)\} \asymp (h_3^*)^4 = n^{-\frac{12}{2d+11}}$ and $\arg\min\{r_2(h) + r_3(h)\} \asymp h_3^* = n^{-\frac{3}{2d+11}}$.

Observe now that $\min\{r_2(h)+r_3(h)\} = O(\min r_1(h))$ is equivalent to $n^{-\frac{12}{2d+11}} = O\left(n^{-4/5}\right)$ which holds if and only if $d \leq 2$. Hence assume that $d \leq 2$. Since $n^{-\frac{12}{2d+11}} = O\left(n^{-4/5}\right)$ also gives $\arg\min\{r_2(h)+r_3(h)\} \asymp h_3^* = O(h_1^*)$, we have

$$\min F(b_1) \asymp n^{-4/5} \text{ and } \arg\min F(b_1) \asymp n^{-1/5}.$$

The case $d > 2$ is symmetric with

$$\min F(b_1) \asymp n^{-\frac{12}{2d+11}} \text{ and } \arg\min F(b_1) \asymp n^{-\frac{3}{2d+11}}.$$

This ends the proof of the Theorem. □

Proof of Theorem 3.4

Observe that the Tchebychev inequality gives

$$\sum_{i=1}^{n} \mathbb{1}(X_i \in \mathcal{X}_0) = n\mathbb{P}(X \in \mathcal{X}_0)\left[1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)\right],$$

so that

$$\widetilde{f}_{1n}(\epsilon) = \left[1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)\right] f_n(\epsilon),$$

where

$$f_n(\epsilon) = \frac{1}{nb_1\mathbb{P}(X \in \mathcal{X}_0)} \sum_{i=1}^{n} \mathbb{1}(X_i \in \mathcal{X}_0) K_1\left(\frac{\varepsilon_i - \epsilon}{b_1}\right).$$

Therefore

$$\widehat{f}_{1n}(\epsilon) - \mathbb{E}f_n(\epsilon) = (f_n(\epsilon) - \mathbb{E}f_n(\epsilon)) + \left(\widehat{f}_{1n}(\epsilon) - \widetilde{f}_{1n}(\epsilon)\right) + O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right) f_n(\epsilon). \tag{3.6.14}$$

Let now $f_{in}(\epsilon)$ be as in Lemma 3.5, and note that $f_n(\epsilon) = (1/n)\sum_{i=1}^{n} f_{in}(\epsilon)$. The second and the third claims in Lemma 3.5 yield, since b_1 goes to 0 under (A_{10}),

$$\frac{\sum_{i=1}^{n} \mathbb{E}\left|f_{in}(\epsilon) - \mathbb{E}f_{in}(\epsilon)\right|^3}{\left(\sum_{i=1}^{n} \mathrm{Var} f_{in}(\epsilon)\right)^3} \leq \frac{\frac{Cnf(\epsilon)}{\mathbb{P}(X\in\mathcal{X}_0)^2 b_1^2}\int |K_1(v)|^3 dv + o\left(\frac{n}{b_1^2}\right)}{\left(\frac{nf(\epsilon)}{\mathbb{P}(X\in\mathcal{X}_0)b_1}\int K_1^2(v)dv + o\left(\frac{n}{b_1}\right)\right)^3} = O(b_1) = o(1).$$

Hence the Lyapounov Central Limit Theorem gives, since nb_1 diverges under (A_{10}),

$$\frac{f_n(\epsilon) - \mathbb{E}f_n(\epsilon)}{\sqrt{\mathrm{Var} f_n(\epsilon)}} = \frac{f_n(\epsilon) - \mathbb{E}f_n(\epsilon)}{\sqrt{\frac{\mathrm{Var} f_{in}(\epsilon)}{n}}} \xrightarrow{d} \mathcal{N}(0,1),$$

which yields, using the second equality in Lemma 3.5,

$$\sqrt{nb_1}\left(f_n(\epsilon) - \mathbb{E}f_n(\epsilon)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{f(\epsilon)}{\mathbb{P}(X \in \mathcal{X}_0)}\int K_1^2(v)dv\right). \tag{3.6.15}$$

Moreover, note that for $nb_0^d b_1^3 \to \infty$ and $nb_0^{2d} \to \infty$,

$$\frac{1}{nb_1^5}\left(\frac{1}{nb_0^d}\right)^2 + \left(\frac{1}{b_1^2} + \frac{b_0^d}{b_1^7}\right)^2\left(\frac{1}{nb_0^d}\right)^3 = O\left(\frac{1}{n^2 b_0^d b_1^3}\right).$$

Therefore, since by Assumptions (A$_{11}$) and (A_9), we have $b_0^4 = O\left(1/(nb_0^d)\right)$, $nb_0^d b_1^3 \to \infty$ and that $nb_0^{2d} \to \infty$, the equality above and (3.6.9) then give

$$\begin{aligned}\widehat{f}_{1n}(\epsilon) - \widetilde{f}_{1n}(\epsilon) &\asymp O_{\mathbb{P}}\left[b_0^4 + \frac{1}{n} + \frac{1}{n^2 b_0^d b_1^3} + \left(\frac{1}{nb_1^5} + \frac{b_0^d}{b_1^3}\right)\left(\frac{1}{nb_0^d}\right)^2 + \left(\frac{1}{b_1^2} + \frac{b_0^d}{b_1^7}\right)\left(\frac{1}{nb_0^d}\right)^3\right]^{1/2} \\ &\asymp O_{\mathbb{P}}\left(b_0^4 + \frac{1}{n} + \frac{1}{n^2 b_0^d b_1^3}\right)^{1/2}.\end{aligned}$$

Hence for b_1 going to 0, we have

$$\sqrt{nb_1}\left(\widehat{f}_{1n}(\epsilon) - \widetilde{f}_{1n}(\epsilon)\right) = O_{\mathbb{P}}\left[nb_1\left(b_0^4 + \frac{1}{n} + \frac{1}{n^2 b_0^d b_1^3}\right)\right]^{1/2} = o_{\mathbb{P}}(1),$$

since $nb_0^4 b_1 = o(1)$ and that $nb_0^d b_1^2 \to \infty$ under Assumption (A$_{11}$). Combining the above result with (3.6.15) and (3.6.14), we obtain

$$\sqrt{nb_1}\left(\widehat{f}_{1n}(\epsilon) - \mathbb{E}f_n(\epsilon)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{f(\epsilon)}{\mathbb{P}\left(X \in \mathcal{X}_0\right)} \int K_1^2(v)dv\right).$$

This ends the proof the Theorem, since the first result of Lemma 3.5 gives

$$\mathbb{E}f_n(\epsilon) = \mathbb{E}f_{1n}(\epsilon) = f(\epsilon) + \frac{b_1^2}{2} f^{(2)}(\epsilon) \int v^2 K_1(v)dv + o\left(b_1^2\right) := \overline{f}_{1n}(\epsilon).\square$$

Appendix A : Proof of the intermediate results

Proof of Lemma 3.1

First note that by (A_7), we have $\int zK_0(z)dz = 0$ and $\int K_0(z)dz = 1$. Therefore, since K_0 is continuous and has a compact support, (A_1), (A_2) and a second-order Taylor expansion, yield, for b_0 small enough and any x in $\mathcal{X}_0$,

$$\begin{aligned}
|\overline{g}_n(x) - g(x)| &= \left| \frac{1}{b_0^d} \int K_0\left(\frac{z-x}{b_0}\right) g(z)dz - g(x) \right| = \left| \int K_0(z) \left[g(x+b_0 z) - g(x) \right] dz \right| \\
&= \left| \int K_0(z) \left[b_0 g^{(1)}(x) z + \frac{b_0^2}{2} z g^{(2)}(x+\theta b_0 z) z^\top \right] dz \right|, \ \theta = \theta(x, b_0 z) \in [0,1] \\
&= \left| b_0 g^{(1)}(x) \int zK_0(z)dz + \frac{b_0^2}{2} \int z g^{(2)}(x+\theta b_0 z) z^\top K_0(z) dz \right| \\
&= \frac{b_0^2}{2} \left| \int z g^{(2)}(x+\theta b_0 z) z^\top K_0(z) dz \right| \leq C b_0^2,
\end{aligned}$$

so that

$$\sup_{x \in \mathcal{X}_0} |\overline{g}_n(x) - g(x)| = O\left(b_0^2\right).$$

This gives the first equality of the lemma. To prove the two last equalities in the Lemma, note that it is sufficient to show that

$$\sup_{x \in \mathcal{X}_0} |\widehat{g}_n(x) - \overline{g}_n(x)| = O_{\mathbb{P}}\left(\frac{\ln n}{n b_0^d}\right)^{1/2},$$

since $\bar{g}_n(x)$ is asymptotically bounded away from 0 over $\mathcal{X}_0$ and that $|\overline{g}_n(x) - g(x)| = O(b_0^2)$ uniformly for x in $\mathcal{X}_0$. This follows from Theorem 1 in Einmahl and Mason (2005). □

Proof of Lemma 3.2

For the first equality in the lemma, set

$$\widehat{r}_n(x) = \frac{1}{n b_0^d} \sum_{j=1}^n Y_j K_0\left(\frac{X_j - x}{b_0}\right), \quad \overline{r}_n(x) = \mathbb{E}\left[\widehat{r}_n(x)\right],$$

and observe that

$$\sup_{x \in \mathcal{X}_0} |\widehat{m}_n(x) - m(x)| \leq \sup_{x \in \mathcal{X}_0} \left| \widehat{m}_n(x) - \frac{\overline{r}_n(x)}{\overline{g}_n(x)} \right| + \sup_{x \in \mathcal{X}_0} \frac{1}{|\overline{g}_n(x)|} |\overline{r}_n(x) - \overline{g}_n(x) m(x)|. \tag{A.1}$$

Consider the first term of (A.1). Note that $\mathbb{E}^{1/4}\left[Y^4 | X = x\right] \leq |m(x)| + \mathbb{E}^{1/4}\left[\varepsilon^4\right]$. The compactness of $\mathcal{X}$ from (A_1), the continuity of $m(\cdot)$ from (A_3) and (A_4) then give that $\mathbb{E}\left[Y^4 | X = x\right] < \infty$ uniformly for $x \in \mathcal{X}_0$. Hence under (A_9), Theorem 2 in Einmahl and Mason (2005) gives

$$\sup_{x \in \mathcal{X}_0} \left| \widehat{m}_n(x) - \frac{\overline{r}_n(x)}{\overline{g}_n(x)} \right| = O_{\mathbb{P}}\left(\frac{\ln n}{n b_0^d}\right)^{1/2}.$$

For the second term in (A.1), a second-order Taylor expansion gives, as in the proof of Lemma 3.1,

$$\sup_{x \in \mathcal{X}_0} |\overline{r}_n(x) - \overline{g}_n(x) m(x)| = O(b_0^2).$$

This gives the result of lemma since Lemma 3.1 implies that $\overline{g}_n(x)$ is bounded away from 0 over $\mathcal{X}_0$ uniformly in x and for b_0 small enough. □

Proof of Lemma 3.3

Note that under (A_8), the Taylor expansion with integral remainder gives, for any $x \in \mathcal{X}_0$ and any integer $i \in [1, n]$,

$$K_1\left(\frac{Y_i - \widehat{m}_n(x) - \epsilon}{h_1}\right) = K_1\left(\frac{Y_i - m(x) - \epsilon}{h_1}\right) - \frac{1}{h_1}\left(\widehat{m}_n(x) - m(x)\right)\int_0^1 K_1^{(1)}\left(\frac{Y_i - \theta_n(x,t)}{h_1}\right)dt,$$

where $\theta_n(x,t) = m(x) + \epsilon + t\left(\widehat{m}_n(x) - m(x)\right)$. Therefore

$$\widetilde{f}_n(\epsilon|x) = f_n(\epsilon|x) - \frac{\widehat{m}_n(x) - m(x)}{\widetilde{g}_n(x)}\left[\frac{1}{nh_0^d h_1^2}\sum_{i=1}^n K_0\left(\frac{X_i - x}{h_0}\right)\int_0^1 K_1^{(1)}\left(\frac{Y_i - \theta_n(x,t)}{h_1}\right)dt\right]. \tag{A.2}$$

Now, observe that if $X_i = z$ and $y \in \mathbb{R}$, the change of variable $e = y - m(z) + h_1 v$ gives, under $(A_1) - (A_5)$ and (A_7),

$$\begin{aligned}
\mathbb{E}_n\left|K_1^{(1)}\left(\frac{Y_i - y}{h_1}\right)\right| &= \mathbb{E}\left|K_1^{(1)}\left(\frac{\varepsilon_i + m(z) - y}{h_1}\right)\right| \\
&= \int\left|K_1^{(1)}\left(\frac{e + m(z) - y}{h_1}\right)\right| f(e)de \\
&= h_1\int |K_1^{(1)}(v)| f\left((y - m(z) + h_1 v)\right)dv \le Ch_1.
\end{aligned}$$

Hence

$$\sup_{1\le i\le n}\int_0^1 \mathbb{E}_n\left|K_1^{(1)}\left(\frac{Y_i - \theta_n(x,t)}{h_1}\right)\right| dt \le Ch_1.$$

With the help of this result and Lemma 3.1, we have

$$\begin{aligned}
&\mathbb{E}_n\left|\frac{1}{nh_0^d h_1}\sum_{i=1}^n K_0\left(\frac{X_i - x}{h_0}\right)\int_0^1 K_1^{(1)}\left(\frac{Y_i - \theta_n(x,t)}{h_1}\right)dt\right| \\
&\le \frac{1}{nh_0^d h_1}\sum_{i=1}^n\left|K_0\left(\frac{X_i - x}{h_0}\right)\right| \times \sup_{1\le i\le n}\int_0^1 \mathbb{E}_n\left|K_1^{(1)}\left(\frac{Y_i - \theta_n(x,t)}{h_1}\right)\right| dt \\
&\le \frac{C}{nh_0^d}\sum_{i=1}^n\left|K_0\left(\frac{X_i - x}{h_0}\right)\right| = O_{\mathbb{P}}(1),
\end{aligned}$$

so that

$$\frac{1}{nh_0^d h_1^2}\sum_{i=1}^n K_0\left(\frac{X_i - x}{h_0}\right)\int_0^1 K_1^{(1)}\left(\frac{Y_i - \theta_n(x,t)}{h_1}\right)dt = O_{\mathbb{P}}\left(\frac{1}{h_1}\right).$$

Hence from (A.2), (3.6.8), Lemma 3.2 and Assumption $(\mathbf{A_0})$, we deduce

$$\widetilde{f}_n(\epsilon|x) = f_n(\epsilon|x) + O_{\mathbb{P}}\left(\frac{1}{h_1}\right)\left(b_0^4 + \frac{\ln n}{nb_0^d}\right)^{1/2} = f_n(\epsilon|x) + o\left(\frac{1}{nh_0^d h_1}\right)^{1/2}.\square$$

Proof of Lemma 3.4 and Lemma 3.5

We just give the proof of Lemma 3.4, the proof of Lemma 3.5 being very similar. For the first equality of Lemma 3.4, note that

$$\begin{aligned}
\mathbb{E}\left[\widetilde{\varphi}_{in}(x,y)\right] &= \frac{1}{h_0^d h_1}\int\int K_0\left(\frac{x_1 - x}{h_0}\right)K_1\left(\frac{y_1 - y}{h_1}\right)\varphi(x_1, y_1)dx_1 dy_1 \\
&= \int\int K_0(z)K_1(v)\varphi\left(x + h_0 z, y + h_1 v\right)dzdv.
\end{aligned}$$

A second-order Taylor expansion gives under (A_6), for z in the support of K_0, v in the support of K_1, and h_0, h_1 small enough,

$$\begin{aligned}
&\varphi\left(x+h_0 z, y+h_1 v\right)-\varphi(x, y) \\
&= h_0 \frac{\partial \varphi(x, y)}{\partial x} z^{\top}+h_1 \frac{\partial \varphi(x, y)}{\partial y} v \\
&\quad+\frac{h_0^2}{2} z \frac{\partial^2 \varphi(x+\theta h_0 z, y+\theta h_1 v)}{\partial^2 x} z^{\top}+h_1 h_0 v \frac{\partial^2 \varphi(x+\theta h_0 z, y+\theta h_1 v)}{\partial x \partial y} z^{\top} \\
&\quad+\frac{h_1^2}{2} \frac{\partial^2 \varphi(x+\theta h_0 z, y+\theta h_1 v)}{\partial^2 y} v^2,
\end{aligned}$$

for some $\theta=\theta(x, y, h_0 z, h_1 v)$ in $[0,1]$. This gives, since $\int K_0(z) d z=\int K_1(v) d v=1$, $\int z K_0(z) d z$ and $\int v K_1(v) d v$ vanish under $(A_7)-(A_8)$, and by the Lebesgue Dominated Convergence Theorem,

$$\begin{aligned}
&\mathbb{E}\left[\widetilde{\varphi}_{in}(x, y)\right]-\varphi(x, y)-\frac{h_0^2}{2} \frac{\partial^2 \varphi(x, y)}{\partial^2 x} \int z K_0(z) z^{\top} d z-\frac{h_1^2}{2} \frac{\partial^2 \varphi(x, y)}{\partial^2 y} \int v^2 K_1(v) d v \\
&= \frac{h_0^2}{2} \int \int z\left(\frac{\partial^2 \varphi(x+\theta h_0 z, y+\theta h_1 v)}{\partial^2 x}-\frac{\partial^2 \varphi(x, y)}{\partial^2 x}\right) z^{\top} K_0(z) K_1(v) d z d v \\
&\quad+h_1 h_0 \int \int v\left(\frac{\partial^2 \varphi(x+\theta h_0 z, y+\theta h_1 v)}{\partial x \partial y}-\frac{\partial^2 \varphi(x, y)}{\partial x \partial y}\right) z^{\top} K_0(z) K_1(v) d z d v \\
&\quad+\frac{h_1^2}{2} \int \int\left(\frac{\partial^2 \varphi(x+\theta h_0 z, y+\theta h_1 v)}{\partial^2 y}-\frac{\partial^2 \varphi(x, y)}{\partial^2 y}\right) v^2 K_0(z) K_1(v) d z d v \\
&= o(h_0^2+h_1^2).
\end{aligned}$$

This proves the first equality of Lemma 3.4. The second equality in Lemma follows similarly, since

$$\begin{aligned}
&\operatorname{Var}[\widetilde{\varphi}_{in}(x, y)]=\mathbb{E}\left[\widetilde{\varphi}_{in}^2(x, y)\right]-\left(\mathbb{E}\left[\widetilde{\varphi}_{in}(x, y)\right]\right)^2 \\
&= \frac{1}{h_0^d h_1} \int \int \varphi\left(x+h_0 z, y+h_1 v\right) K_0^2(z) K_1^2(v) d z d v+O(1) \\
&= \frac{\varphi(x, y)}{h_0^d h_1} \int \int K_0^2(z) K_1^2(v) d z d v+o\left(\frac{1}{h_0^d h_1}\right).
\end{aligned}$$

The last statement of Lemma 3.4 is immediate, since the Triangular and Convex inequalities give

$$\begin{aligned}
\mathbb{E}\left|\widetilde{\varphi}_{in}(x, y)-\mathbb{E} \widetilde{\varphi}_{in}(x, y)\right|^3 &\leq C \mathbb{E}\left|\widetilde{\varphi}_{in}(x, y)\right|^3 \\
&\leq \frac{C \varphi(x, y)}{h_0^{2d} h_1^2} \int \int\left|K_0(z) K_1(v)\right|^3 d z d v+o\left(\frac{1}{h_0^{2d} h_1^2}\right) . \square
\end{aligned}$$

Proof of Lemma 3.6

The order of S_n follows from Lemma 3.8 and Lemma 3.9. In fact, since

$$\begin{aligned}
\mathbb{1}(X_i \in \mathcal{X}_0)\left(\widehat{m}_{in}-m(X_i)\right) &= \frac{\mathbb{1}(X_i \in \mathcal{X}_0)}{n b_0^d \widehat{g}_{in}} \sum_{j=1, j \neq i}^n\left(m(X_j)+\varepsilon_j-m(X_i)\right) K_0\left(\frac{X_j-X_i}{b_0}\right) \\
&= \beta_{in}+\Sigma_{in},
\end{aligned}$$

Lemma 3.8 and Lemma 3.9 give

$$S_n=O_{\mathbb{P}}\left[b_0^2\left(n b_1^2+(n b_1)^{1/2}\right)+\left(n b_1^4+\frac{b_1}{b_0^d}\right)^{1/2}\right],$$

which gives the result for S_n.

For T_n, define for any $1 \leq i \leq n$,

$$\mathbb{E}_{in}[\cdot] = \mathbb{E}_n\left[X_1, \ldots, X_n, \varepsilon_k, k \neq i\right].$$

Therefore, since $(\widehat{m}_{in} - m(X_i))$ depends only upon $(X_1, \ldots, X_n, \varepsilon_k, k \neq i)$, we have

$$\begin{aligned}
\mathbb{E}_n[T_n] &= \mathbb{E}_n\left[\sum_{i=1}^n \mathbb{E}_{in}\left[\mathbb{1}\left(X_i \in \mathcal{X}_0\right)\left(\widehat{m}_{in} - m(X_i)\right)^2 K_1^{(2)}\left(\frac{\varepsilon_i - \epsilon}{b_1}\right)\right]\right] \\
&= \mathbb{E}_n\left[\sum_{i=1}^n \mathbb{1}\left(X_i \in \mathcal{X}_0\right)\left(\widehat{m}_{in} - m(X_i)\right)^2 \mathbb{E}_{in}\left[K_1^{(2)}\left(\frac{\varepsilon_i - \epsilon}{b_1}\right)\right]\right],
\end{aligned}$$

with, using (A_4) and Lemma 3.7-(3.6.2),

$$\left|\mathbb{E}_{in}\left[K_1^{(2)}\left(\frac{\varepsilon_i - \epsilon}{b_1}\right)\right]\right| = \left|\int K_1^{(2)}\left(\frac{e - \epsilon}{b_1}\right) f(e) de\right| \leq C b_1^3.$$

Hence this bound, the equality above, the Cauchy-Schwarz inequality and Lemma 3.10 yield that

$$\begin{aligned}
\left|\mathbb{E}_n\left[T_n\right]\right| &\leq C b_1^3 \sum_{i=1}^n \mathbb{E}_n\left[\mathbb{1}\left(X_i \in \mathcal{X}_0\right)\left(\widehat{m}_{in} - m(X_i)\right)^2\right] \\
&\leq C n b_1^3 \left(\sup_{1 \leq i \leq n} \mathbb{E}_n\left[\mathbb{1}\left(X_i \in \mathcal{X}_0\right)\left(\widehat{m}_{in} - m(X_i)\right)^4\right]\right)^{1/2} \\
&\leq O_{\mathbb{P}}\left(n b_1^3\right)\left(b_0^4 + \frac{1}{n b_0^d}\right). \qquad \text{(A.3)}
\end{aligned}$$

For the conditional variance of T_n, Lemma 3.12 gives

$$\begin{aligned}
\operatorname{Var}_n(T_n) &= \sum_{i=1}^n \operatorname{Var}_n\left(\zeta_{in}\right) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \operatorname{Cov}_n\left(\zeta_{in}, \zeta_{jn}\right) \\
&= O_{\mathbb{P}}\left(n b_1\right)\left(b_0^4 + \frac{b_1}{n b_0^d}\right)^2 + O_{\mathbb{P}}\left(n^2 b_0^d b_1^{7/2}\right)\left(b_0^4 + \frac{1}{n b_0^d}\right)^2.
\end{aligned}$$

Therefore, since b_1 goes to 0 under (A_{10}), this order and (A.3) yield, applying the Tchebychev inequality,

$$\begin{aligned}
T_n &= O_{\mathbb{P}}\left[\left(n b_1^3\right)\left(b_0^4 + \frac{1}{n b_0^d}\right) + \left(n b_1\right)^{1/2}\left(b_0^4 + \frac{b_1}{n b_0^d}\right) + \left(n^2 b_0^d b_1^{7/2}\right)^{1/2}\left(b_0^4 + \frac{1}{n b_0^d}\right)\right] \\
&= O_{\mathbb{P}}\left[\left(n b_1^3 + \left(n b_1\right)^{1/2} + \left(n^2 b_0^d b_1^3\right)^{1/2}\right)\left(b_0^4 + \frac{1}{n b_0^d}\right)\right].
\end{aligned}$$

which gives the result for T_n.

We now compute the order of R_n. For this, define

$$\begin{aligned}
I_{in} &= \int_0^1 (1 - t)^2 K_1^{(3)}\left(\frac{\varepsilon_i - t(\widehat{m}_{in} - m(X_i)) - \epsilon}{b_1}\right) dt, \\
R_{in} &= \mathbb{1}\left(X_i \in \mathcal{X}_0\right)\left(\widehat{m}_{in} - m(X_i)\right)^3 I_{in},
\end{aligned}$$

and note that $R_n = \sum_{i=1}^{n} R_{in}$. The order of R_n is derived by computing its conditional mean and its conditional variance. For the conditional mean, observe that

$$\begin{aligned}\mathbb{E}_n[R_n] &= \mathbb{E}_n\left[\sum_{i=1}^{n}\mathbb{E}_{in}[R_{in}]\right]\\ &= \mathbb{E}_n\left[\sum_{i=1}^{n}\mathbb{1}(X_i\in\mathcal{X}_0)(\widehat{m}_{in}-m(X_i))^3\mathbb{E}_{in}[I_{in}]\right],\end{aligned}$$

with, using (A_4) and Lemma 3.7-(3.6.3),

$$\begin{aligned}|\mathbb{E}_{in}[I_{in}]| &= \left|\int_0^1(1-t)^2\left[\int K_1^{(3)}\left(\frac{e-t(\widehat{m}_{in}-m(X_i))-\epsilon}{b_1}\right)f(e)de\right]dt\right|\\ &\leq Cb_1^3.\end{aligned}$$

Therefore the Holder inequality and Lemma 3.10 yield

$$\begin{aligned}|\mathbb{E}_n[R_n]| &\leq Cb_1^3\sum_{i=1}^{n}\mathbb{E}_n\left[\left|\mathbb{1}(X_i\in\mathcal{X}_0)(\widehat{m}_{in}-m(X_i))\right|^3\right]\\ &\leq Cb_1^3\sum_{i=1}^{n}\mathbb{E}_n^{3/4}\left[\mathbb{1}(X_i\in\mathcal{X}_0)(\widehat{m}_{in}-m(X_i))^4\right]\\ &\leq O_{\mathbb{P}}\left(nb_1^3\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^{3/2}. \end{aligned}\tag{A.4}$$

For the conditional covariance of R_n, note that Lemma 3.11 allows to write

$$\mathrm{Var}_n(R_n) = \sum_{i=1}^{n}\mathrm{Var}_n(R_{in}) + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\Big(\|X_i-X_j\|\leq Cb_0\Big)\mathrm{Cov}_n(R_{in},R_{jn}),\tag{A.5}$$

and consider the first term in (A.5). We have

$$\mathrm{Var}_n(R_{in}) \leq \mathbb{E}_n\left[R_{in}^2\right] \leq \mathbb{E}_n\Big[\mathbb{1}(X_i\in\mathcal{X}_0)(\widehat{m}_{in}-m(X_i))^6\mathbb{E}_{in}\left[I_{in}^2\right]\Big],$$

with, using (A_4), the Cauchy-Schwarz inequality and Lemma 3.7-(3.6.3),

$$\begin{aligned}\mathbb{E}_{in}\left[I_{in}^2\right] &\leq C\mathbb{E}_{in}\left[\int_0^1 K_1^{(3)}\left(\frac{\varepsilon_i-t(\widehat{m}_{in}-m(X_i))-\epsilon}{b_1}\right)^2dt\right]\\ &\leq C\int_0^1\left[\int K_1^{(3)}\left(\frac{e-t(\widehat{m}_{in}-m(X_i))-\epsilon}{b_1}\right)^2 f(e)de\right]dt\\ &\leq Cb_1,\end{aligned}$$

so that

$$\mathrm{Var}_n(R_{in}) \leq Cb_1\mathbb{E}_n\left[\mathbb{1}(X_i\in\mathcal{X}_0)(\widehat{m}_{in}-m(X_i))^6\right].$$

Therefore form Lemma 3.10, we deduce

$$\begin{aligned}\sum_{i=1}^{n}\mathrm{Var}_n(R_{in}) &\leq Cnb_1\sup_{1\leq i\leq n}\mathbb{E}_n\left[\mathbb{1}(X_i\in\mathcal{X}_0)(\widehat{m}_{in}-m(X_i))^6\right]\\ &\leq O_{\mathbb{P}}(nb_1)\left(b_0^4+\frac{1}{nb_0^d}\right)^3. \end{aligned}\tag{A.6}$$

For the second term in (A.5), the Cauchy-Schwarz inequality gives, with the help of the above result for $\mathrm{Var}_n(R_{in})$,

$$\begin{aligned}|\mathrm{Cov}_n(R_{in}, R_{jn})| &\leq (\mathrm{Var}_n(R_{in})\,\mathrm{Var}_n(R_{jn}))^{1/2}\\ &\leq Cb_1 \sup_{1\leq i\leq n} \mathbb{E}_n\left[\mathbb{1}(X_i \in \mathcal{X}_0)(\widehat{m}_{in} - m(X_i))^6\right].\end{aligned}$$

Hence by Lemma 3.10 and the Markov inequality, we have

$$\begin{aligned}&\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(\|X_i - X_j\| \leq Cb_0\right)|\mathrm{Cov}_n(R_{in}, R_{jn})|\\ &\leq O_{\mathbb{P}}(b_1)\left(b_0^4 + \frac{1}{nb_0^d}\right)^3 \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(\|X_i - X_j\| \leq Cb_0\right)\\ &\leq O_{\mathbb{P}}(b_1)\left(b_0^4 + \frac{1}{nb_0^d}\right)^3 \left(n^2 b_0^d\right).\end{aligned}$$

This order, (A.6) and (A.5) give, since nb_0^d diverges under (A_9),

$$\mathrm{Var}(R_n) = O_{\mathbb{P}}\left(b_0^4 + \frac{1}{nb_0^d}\right)^3 \left(n^2 b_0^d b_1\right).$$

Finally, with the help of this result, (A.4) and the Tchebychev inequality, we arrive at

$$\begin{aligned}R_n &= O_{\mathbb{P}}\left[\left(nb_1^3\right)\left(b_0^4 + \frac{1}{nb_0^d}\right)^{3/2} + \left(n^2 b_0^d b_1\right)^{1/2}\left(b_0^4 + \frac{1}{nb_0^d}\right)^{3/2}\right]\\ &= O_{\mathbb{P}}\left[\left(nb_1^3 + \left(n^2 b_0^d b_1\right)^{1/2}\right)\left(b_0^4 + \frac{1}{nb_0^d}\right)^{3/2}\right]. \square\end{aligned}$$

Proof of Lemma 3.7

Set $h_p(e) = e^p f(e)$, $p \in [0, 2]$. For the first inequality of (3.6.1), note that under (A_5) and (A_8), the change of variable $e = \epsilon + b_1 v$ give, for any integer $\ell \in [1, 3]$,

$$\begin{aligned}\left|\int K_1^{(\ell)}\left(\frac{e-\epsilon}{b_1}\right)^2 e^p f(e)de\right| &= \left|b_1 \int K_1^{(\ell)}(v)^2 h_p(\epsilon + b_1 v)dv\right|\\ &\leq b_1 \sup_{t\in\mathbb{R}}|h_p(t)| \int |K_1^{(\ell)}(v)^2|dv\\ &\leq Cb_1, \qquad \text{(A.7)}\end{aligned}$$

which yields the first inequality in (3.6.1). For the second inequality in (3.6.1), observe that $f(\cdot)$ has a bounded continuous derivative under (A_5), and that $\int K_1^{(\ell)}(v)dv = 0$ under (A_8). Therefore, since $h_p(\cdot)$ has bounded second order derivatives under (A_7), the Taylor inequality yields that

$$\begin{aligned}\left|\int K_1^{(\ell)}\left(\frac{e-\epsilon}{b_1}\right) e^p f(e)de\right| &= b_1\left|\int K_1^{(\ell)}(v)\left[h_p(\epsilon + b_1 v) - h_p(\epsilon)\right]\right| dv\\ &\leq b_1^2 \sup_{t\in\mathbb{R}}|h_p^{(1)}(t)| \int |vK_1^{(\ell)}(v)|dv \leq Cb_1^2.\end{aligned}$$

which completes the proof of (3.6.1).

The first inequalities of (3.6.2) and (3.6.3) follow directly from (A.7). The second bounds in (3.6.2) and (3.6.3) are proved simultaneously. For this, note that for any integer $\ell \in \{2,3\}$,

$$\int K_1^{(\ell)}\left(\frac{e-\epsilon}{b_1}\right) h_p(e)de = b_1 \int K_1^{(\ell)}(v) h_p(\epsilon + b_1 v)dv.$$

Under (A_8), $K_1(\cdot)$ is symmetric, has a compact support and two continuous derivatives, with $\int K_1^{(\ell)}(v)dv = 0$ and $\int vK_1^{(\ell)}(v)dv = 0$ for $\ell \in \{2,3\}$. Hence, since by (A_5) h_p has bounded continuous second order derivatives, this gives for some $\theta = \theta(\epsilon, b_1 v)$,

$$\begin{aligned}
\left|\int K_1^{(\ell)}\left(\frac{e-\epsilon}{b_1}\right) h_p(e)de\right| &= \left|b_1 \int K_1^{(\ell)}(v)\left[h_p(\epsilon+b_1v) - h_p(\epsilon)\right]dv\right| \\
&= \left|b_1 \int K_1^{(\ell)}(v)\left[b_1 v h_p^{(1)}(\epsilon) + \frac{b_1^2 v^2}{2} h_p^{(2)}(\epsilon + \theta b_1 v)\right]dv\right| \\
&= \left|\frac{b_1^3}{2}\int v^2 K_1^{(\ell)}(v) h_p^{(2)}(\epsilon+\theta b_1 v)dv\right| \\
&\leq \frac{b_1^3}{2}\sup_{t\in\mathbb{R}} |h_p^{(2)}(t)| \int \left|v^2 K_1^{(\ell)}(v)\right| dv \leq C b_1^3. \square
\end{aligned}$$

Proof of Lemma 3.8

Assumption (A_4) and Lemma 3.7-(3.6.1) give

$$\begin{aligned}
\left|\mathbb{E}_n\left[\sum_{i=1}^n \beta_{in} K_1^{(1)}\left(\frac{\varepsilon_i - \epsilon}{b_1}\right)\right]\right| &= \left|\mathbb{E}\left[K_1^{(1)}\left(\frac{\varepsilon-\epsilon}{b_1}\right)\right]\sum_{i=1}^n \beta_{in}\right| \leq Cnb_1^2 \max_{1\leq i\leq n}|\beta_{in}|, \\
\mathrm{Var}_n\left[\sum_{i=1}^n \beta_{in} K_1^{(1)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right] &\leq \sum_{i=1}^n \beta_{in}^2 \mathbb{E}\left[K_1^{(1)}\left(\frac{\varepsilon-\epsilon}{b_1}\right)^2\right] \leq Cnb_1 \max_{1\leq i\leq n}|\beta_{in}|^2.
\end{aligned}$$

Hence the (conditional) Markov inequality gives

$$\sum_{i=1}^n \beta_{in} K_1^{(1)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right) = O_{\mathbb{P}}\left(nb_1^2 + (nb_1)^{1/2}\right)\max_{1\leq i\leq n}|\beta_{in}|,$$

so that the lemma follows if we can prove that

$$\sup_{1\leq i\leq n}|\beta_{in}| = O_{\mathbb{P}}\left(b_0^2\right), \tag{A.8}$$

as established now. For this, define

$$\zeta_j(x) = \mathbb{1}\left(x\in\mathcal{X}_0\right)\left(m(X_j)-m(x)\right)K_0\left(\frac{X_j - x}{b_0}\right), \quad \nu_{in}(x) = \frac{1}{(n-1)b_0^d}\sum_{j=1, j\neq i}^n \left(\zeta_j(x) - \mathbb{E}[\zeta_j(x)]\right),$$

and $\bar{\nu}_n(x) = \mathbb{E}[\zeta_j(x)]/b_0^d$, so that

$$\beta_{in} = \frac{n-1}{n}\frac{\nu_{in}(X_i)+\bar{\nu}_n(X_i)}{\widehat{g}_{in}}.$$

For $\max_{1\leq i\leq n}|\bar{\nu}_n(X_i)|$, first observe that a second-order Taylor expansion applied successively to $g(\cdot)$ and $m(\cdot)$ give, for b_0 small enough, and for any x, z in $\mathcal{X}$,

$$\begin{aligned}
&[m(x+b_0z) - m(x)]\, g(x+b_0z) \\
&= \left[b_0 m^{(1)}(x)z + \frac{b_0^2}{2} z m^{(2)}(x+\zeta_1 b_0 z)z^\top\right]\left[g(x) + b_0 g^{(1)}(x)z + \frac{b_0^2}{2} z g^{(2)}(x+\zeta_2 b_0 z)z^\top\right],
\end{aligned}$$

for some $\zeta_1 = \zeta_1(x, b_0 z)$ and $\zeta_2 = \zeta_2(x, b_0 z)$ in $[0,1]$. Therefore, since $\int zK(z)dz = 0$ under (A_7), it follows that, by (A_1), (A_2) and (A_3),

$$\begin{aligned}\max_{1\le i\le n} |\bar{\nu}_n(X_i)| &\le \sup_{x\in\mathcal{X}_0} |\bar{\nu}_n(x)| = \sup_{x\in\mathcal{X}_0}\left|\int (m(x+b_0z) - m(x))\, K_0(z) g(x+b_0z)dz\right| \\ &\le Cb_0^2. \end{aligned} \tag{A.9}$$

Consider now the term $\max_{1\le i\le n}|\nu_{in}(X_i)|$. The Bernstein inequality (see e.g. Serfling (2002)) and (A_4) give, for any $t > 0$,

$$\begin{aligned}\mathbb{P}\left(\max_{1\le i\le n}|\nu_{in}(X_i)| \ge t\right) &\le \sum_{i=1}^n \mathbb{P}\left(|\nu_{in}(X_i)| \ge t\right) \le \sum_{i=1}^n \int \mathbb{P}\left(|\nu_{in}(x)| \ge t\,|X_i = x\right) g(x)dx \\ &\le 2n\exp\left(-\frac{(n-1)t^2}{2\sup_{x\in\mathcal{X}_0}\mathrm{Var}(\zeta_j(x)/b_0^d) + \frac{4M}{3b_0^d}t}\right),\end{aligned}$$

where M is such that $\sup_{x\in\mathcal{X}_0}|\zeta_j(x)| \le M$. The definition of $\mathcal{X}_0$ given in (A_2), (A_3), (A_7) and the standard Taylor expansion yield, for b_0 small enough,

$$\sup_{x\in\mathcal{X}_0}|\zeta_j(x)| \le Cb_0, \quad \sup_{x\in\mathcal{X}_0}\mathrm{Var}(\zeta_j(x)/b_0^d) \le \frac{1}{b_0^d}\sup_{x\in\mathcal{X}_0}\int (m(x+b_0z) - m(x))^2 K_0^2(z) g(x+b_0z)dz \le \frac{Cb_0^2}{b_0^d},$$

so that, for any $t \ge 0$,

$$\mathbb{P}\left(\max_{1\le i\le n}|\nu_{in}(X_i)| \ge t\right) \le 2n\exp\left(-\frac{(n-1)b_0^d t^2/b_0^2}{C + Ct/b_0}\right).$$

This gives

$$\mathbb{P}\left(\max_{1\le i\le n}|\nu_{in}(X_i)| \ge \left(\frac{b_0^2\ln n}{(n-1)b_0^d}\right)^{1/2} t\right) \le 2n\exp\left(-\frac{t^2\ln n}{C + Ct\left(\frac{\ln n}{(n-1)b_0^d}\right)^{1/2}}\right) = o(1),$$

provided that t is large enough and under (A_9). It then follows that

$$\max_{1\le i\le n}|\nu_{in}(X_i)| = O_{\mathbb{P}}\left(\frac{b_0^2\ln n}{nb_0^d}\right)^{1/2}.$$

This bound, (A.9) and Lemma 3.1 show that (A.8) is proved, since $b_0^2\ln n/(nb_0^d) = O\left(b_0^4\right)$ under (A_9), and that

$$\beta_{in} = \frac{n-1}{n}\frac{\nu_{in}(X_i) + \bar{\nu}_n(X_i)}{\widehat{g}_{in}}.\square$$

Proof of Lemma 3.9

Note that (A_4) gives that Σ_{in} is independent of ε_i, and that $\mathbb{E}_n[\Sigma_{in}] = 0$. This yields

$$\mathbb{E}_n\left[\sum_{i=1}^n \Sigma_{in} K_1^{(1)}\left(\frac{\varepsilon_i - \epsilon}{b_1}\right)\right] = 0. \tag{A.10}$$

Moreover, observe that

$$\begin{aligned}&\mathrm{Var}_n\left[\sum_{i=1}^n \Sigma_{in}K_1^{(1)}\left(\frac{\varepsilon_i - \epsilon}{b_1}\right)\right] \\ &= \sum_{i=1}^n \mathrm{Var}_n\left[\Sigma_{in}K_1^{(1)}\left(\frac{\varepsilon_i - \epsilon}{b_1}\right)\right] + \sum_{i=1}^n\sum_{\substack{j=1\\ j\ne i}}^n \mathrm{Cov}_n\left[\Sigma_{in}K_1^{(1)}\left(\frac{\varepsilon_i - \epsilon}{b_1}\right), \Sigma_{jn}K_1^{(1)}\left(\frac{\varepsilon_j - \epsilon}{b_1}\right)\right].\end{aligned} \tag{A.11}$$

For the sum of variances in (A.11), Lemma 3.7-(3.6.1) and (A_4) give

$$\begin{aligned}
\sum_{i=1}^{n} \mathrm{Var}_n \left[\Sigma_{in} K_1^{(1)} \left(\frac{\varepsilon_i - \epsilon}{b_1} \right) \right] &\leq \sum_{i=1}^{n} \mathbb{E}_n \left[\Sigma_{in}^2 \right] \mathbb{E} \left[K_1^{(1)} \left(\frac{\varepsilon_i - \epsilon}{b_1} \right)^2 \right] \\
&\leq \frac{C b_1 \sigma^2}{(n b_0^d)^2} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{\mathbb{1}(X_i \in \mathcal{X}_0)}{\widehat{g}_{in}^2} K_0^2 \left(\frac{X_j - X_i}{b_0} \right) \\
&\leq \frac{C b_1 \sigma^2}{n b_0^d} \sum_{i=1}^{n} \frac{\mathbb{1}(X_i \in \mathcal{X}_0) \widetilde{g}_{in}}{\widehat{g}_{in}^2} \,, \qquad \text{(A.12)}
\end{aligned}$$

where $\sigma^2 = \mathrm{Var}(\varepsilon)$ and

$$\widetilde{g}_{in} = \frac{1}{n b_0^d} \sum_{j=1, j \neq i}^{n} K_0^2 \left(\frac{X_j - X_i}{b_0} \right).$$

For the sum of conditional covariances in (A.11), observe that by (A_4) we have

$$\begin{aligned}
&\sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathrm{Cov}_n \left[\Sigma_{in} K_1^{(1)} \left(\frac{\varepsilon_i - \epsilon}{b_1} \right), \Sigma_{jn} K_1^{(1)} \left(\frac{\varepsilon_j - \epsilon}{b_1} \right) \right] \\
&= \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathbb{E}_n \left[\Sigma_{in} \Sigma_{jn} K_1^{(1)} \left(\frac{\varepsilon_i - \epsilon}{b_1} \right) K_1^{(1)} \left(\frac{\varepsilon_j - \epsilon}{b_1} \right) \right] \\
&= \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{\mathbb{1}(X_i \in \mathcal{X}_0) \mathbb{1}(X_j \in \mathcal{X}_0)}{(n b_0^d)^2 \widehat{g}_{in} \widehat{g}_{jn}} \sum_{\substack{k=1 \\ k \neq i}}^{n} \sum_{\substack{\ell=1 \\ \ell \neq j}}^{n} K_0 \left(\frac{X_k - X_i}{b_0} \right) K_0 \left(\frac{X_\ell - X_j}{b_0} \right) \mathbb{E} \left[\xi_{ki} \xi_{\ell j} \right],
\end{aligned}$$

where

$$\xi_{ki} = \varepsilon_k K_1^{(1)} \left(\frac{\varepsilon_i - \epsilon}{b_1} \right).$$

Moreover, under (A_4), it is seen that for $k \neq \ell$, $\mathbb{E}[\xi_{ki} \xi_{\ell j}] = 0$ when $\mathrm{Card}\{i, j, k, \ell\} \geq 3$. Therefore the symmetry of K_0 yields that

$$\begin{aligned}
&\sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathrm{Cov}_n \left[\Sigma_{in} K_1^{(1)} \left(\frac{\varepsilon_i - \epsilon}{b_1} \right), \Sigma_{jn} K_1^{(1)} \left(\frac{\varepsilon_j - \epsilon}{b_1} \right) \right] \\
&= \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{\mathbb{1}(X_i \in \mathcal{X}_0) \mathbb{1}(X_j \in \mathcal{X}_0)}{(n b_0^d)^2 \widehat{g}_{in} \widehat{g}_{jn}} K_0^2 \left(\frac{X_j - X_i}{b_0} \right) \mathbb{E}^2 \left[\varepsilon K_1^{(1)} \left(\frac{\varepsilon - \epsilon}{b_1} \right) \right] \\
&\quad + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{\mathbb{1}(X_i \in \mathcal{X}_0) \mathbb{1}(X_j \in \mathcal{X}_0)}{(n b_0^d)^2 \widehat{g}_{in} \widehat{g}_{jn}} \sum_{\substack{k=1 \\ k \neq i,j}}^{n} K_0 \left(\frac{X_k - X_i}{b_0} \right) K_0 \left(\frac{X_k - X_j}{b_0} \right) \mathbb{E}[\varepsilon^2] \mathbb{E}^2 \left[K_1^{(1)} \left(\frac{\varepsilon - \epsilon}{b_1} \right) \right].
\end{aligned}$$

Therefore, since

$$\sup_{1 \leq j \leq n} \left(\frac{\mathbb{1}(X_j \in \mathcal{X}_0)}{|\widehat{g}_{jn}|} \right) = O_{\mathbb{P}}(1)$$

by Lemma 3.1, Lemma 3.7-(3.6.1) and (A_4) then give

$$\begin{aligned}
&\left| \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathrm{Cov}_n \left[\Sigma_{in} K_1^{(1)} \left(\frac{\varepsilon_i - \epsilon}{b_1} \right), \Sigma_{jn} K_1^{(1)} \left(\frac{\varepsilon_j - \epsilon}{b_1} \right) \right] \right| \\
&= O_{\mathbb{P}} \left(\frac{b_1^4}{n b_0^d} \right) \sum_{i=1}^{n} \frac{\mathbb{1}(X_i \in \mathcal{X}_0) \widetilde{g}_{in}}{|\widehat{g}_{in}|} + O_{\mathbb{P}}(b_1^4) \sum_{i=1}^{n} \frac{\mathbb{1}(X_i \in \mathcal{X}_0) |g_{in}|}{|\widehat{g}_{in}|} \,, \qquad \text{(A.13)}
\end{aligned}$$

where $\widetilde{g}_{in}$ is defined as in (A.12) and

$$g_{in} = \frac{1}{(nb_0^d)^2} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{k=1 \\ k\neq j,i}}^{n} K_0\left(\frac{X_k - X_i}{b_0}\right) K_0\left(\frac{X_k - X_j}{b_0}\right).$$

The order of the first term in (A.13) follows from Lemma 3.1, which gives

$$\sum_{i=1}^{n} \frac{\mathbb{1}(X_i \in \mathcal{X}_0)\widetilde{g}_{in}}{|\widehat{g}_{in}|} = O_{\mathbb{P}}(n). \tag{A.14}$$

Again, by Lemma 3.1, we have

$$\sum_{i=1}^{n} \frac{\mathbb{1}(X_i \in \mathcal{X}_0)|g_{in}|}{|\widehat{g}_{in}|} = O_{\mathbb{P}}(1) \sum_{i=1}^{n} \mathbb{1}\left(X_i \in \mathcal{X}_0\right)|g_{in}|,$$

with, using the changes of variables $x_1 = x_3 + b_0 z_1$, $x_2 = x_3 + b_0 z_2$,

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{n} \mathbb{1}\left(X_i \in \mathcal{X}_0\right)|g_{in}|\right] &\leq \frac{Cn^3}{(nb_0^d)^2} \mathbb{E}\left|K_1\left(\frac{X_3 - X_1}{h}\right) K_1\left(\frac{X_3 - X_2}{h}\right)\right| \\
&= \frac{Cn^3}{n^2h^2} \int_{\mathcal{X}_0^3} \left|K_1\left(\frac{x_3 - x_1}{h}\right) K_1\left(\frac{x_3 - x_2}{h}\right)\right| \prod_{k=1}^{3} g(x_k)dx_k \\
&\leq \frac{Cn^3 b_0^{2d}}{(nb_0^d)^2}.
\end{aligned}$$

These bounds and the equality above, give under (A_2) and (A_7),

$$\sum_{i=1}^{n} \frac{\mathbb{1}(X_i \in \mathcal{X}_0)|g_{in}|}{|\widehat{g}_{in}|} = O_{\mathbb{P}}(n).$$

Hence from (A.14), (A.13), (A.12), (A.11) and Lemma 3.1, we deduce, for b_1 small enough,

$$\begin{aligned}
&\mathrm{Var}_n\left[\sum_{i=1}^{n} \Sigma_{in} K_1^{(1)}\left(\frac{\varepsilon_i - \epsilon}{b_1}\right)\right] \\
&= O_{\mathbb{P}}\left(\frac{b_1}{nb_0^d}\right) \sum_{i=1}^{n} \frac{\mathbb{1}(X_i \in \mathcal{X}_0)\widetilde{g}_{in}}{\widehat{g}_{in}^2} + O_{\mathbb{P}}\left(\frac{b_1^4}{nb_0^d}\right) \sum_{i=1}^{n} \frac{\mathbb{1}(X_i \in \mathcal{X}_0)\widetilde{g}_{in}}{|\widehat{g}_{in}|} + O_{\mathbb{P}}(b_1^4) \sum_{i=1}^{n} \frac{\mathbb{1}(X_i \in \mathcal{X}_0)|g_{in}|}{|\widehat{g}_{in}|} \\
&= O_{\mathbb{P}}\left(\frac{b_1}{b_0^d} + \frac{b_1^4}{b_0^d} + nb_1^4\right) = O_{\mathbb{P}}\left(\frac{b_1}{b_0^d} + nb_1^4\right).
\end{aligned}$$

Finally, this order, (A.10) and the Tchebychev inequality give

$$\sum_{i=1}^{n} \Sigma_{in} K_1^{(1)}\left(\frac{\varepsilon_i - \epsilon}{b_1}\right) = O_{\mathbb{P}}\left(\frac{b_1}{b_0^d} + nb_1^4\right)^{1/2}.\square$$

Proof of Lemma 3.10

Define β_{in} as in Lemma 3.8 and set

$$g_{in} = \frac{1}{nb_0^d} \sum_{j=1, j\neq i}^{n} K_0^4\left(\frac{X_j - X_i}{b_0}\right), \quad \widetilde{g}_{in} = \frac{1}{nb_0^d} \sum_{j=1, j\neq i}^{n} K_0^2\left(\frac{X_j - X_i}{b_0}\right).$$

The proof of the lemma is based on the following bound :

$$\mathbb{E}_n\left[\mathbb{1}\left(X_i \in \mathcal{X}_0\right)\left(\widehat{m}_{in} - m(X_i)\right)^k\right] \leq C\left[\beta_{in}^k + \frac{\mathbb{1}\left(X_i \in \mathcal{X}_0\right)\widetilde{g}_{in}^{k/2}}{(nb_0^d)^{(k/2)}\widehat{g}_{in}^k}\right], \quad k \in \{4, 6\}. \tag{A.15}$$

Indeed, taking successively $k = 4$ and $k = 6$ in (A.15), we have, by (A.8), Lemma 3.1 and (A_9),

$$\sup_{1\leq i\leq n} \mathbb{E}_n\left[\mathbb{1}\left(X_i \in \mathcal{X}_0\right)\left(\widehat{m}_{in} - m(X_i)\right)^4\right] = O_{\mathbb{P}}\left(b_0^8 + \frac{1}{(nb_0^d)^2}\right) = O_{\mathbb{P}}\left(b_0^4 + \frac{1}{nb_0^d}\right)^2,$$
$$\sup_{1\leq i\leq n} \mathbb{E}_n\left[\mathbb{1}\left(X_i \in \mathcal{X}_0\right)\left(\widehat{m}_{in} - m(X_i)\right)^6\right] = O_{\mathbb{P}}\left(b_0^{12} + \frac{1}{(nb_0^d)^3}\right) = O_{\mathbb{P}}\left(b_0^4 + \frac{1}{nb_0^d}\right)^3,$$

which gives the results of the Lemma. Hence it remains to prove (A.15). For this, define β_{in} and Σ_{in} respectively as in Lemma 3.8 and Lemma 3.9. Since $\mathbb{1}(X_i \in \mathcal{X}_0)\left(\widehat{m}_{in} - m(X_i)\right) = \beta_{in} + \Sigma_{in}$, and that β_{in} depends only on $(X_1, \ldots, X_n)$, this gives, for $k \in \{4, 6\}$

$$\mathbb{E}_n\left[\mathbb{1}(X_i \in \mathcal{X}_0)\left(\widehat{m}_{in} - m(X_i)\right)^k\right] \leq C\beta_{in}^k + C\mathbb{E}_n\left[\Sigma_{in}^k\right]. \tag{A.16}$$

The order of the second term of bound (A.16) is computed by applying Theorem 2 in Whittle (1960) or the Marcinkiewicz-Zygmund inequality (see e.g Chow and Teicher, 2003, p. 386). These inequalities show that for linear form $L = \sum_{j=1}^n a_j\zeta_j$ with independent mean-zero random variables $\zeta_1, \ldots, \zeta_n$, it holds that, for any $k \geq 1$,

$$\mathbb{E}\left|L^k\right| \leq C(k)\left[\sum_{j=1}^n a_j^2\mathbb{E}^{2/k}\left|\zeta_j^k\right|\right]^{k/2},$$

where $C(k)$ is a positive real depending only on k. Now, observe that for any $i \in [1, n]$,

$$\Sigma_{in} = \sum_{j=1, j\neq i}^n \sigma_{jin}, \quad \sigma_{jin} = \frac{\mathbb{1}\left(X_i \in \mathcal{X}_0\right)}{nb_0^d\widehat{g}_{in}}\varepsilon_j K_0\left(\frac{X_j - X_i}{b_0}\right).$$

Since under (A_4), the σ_{jin}'s, $j \in [1, n]$, are centered independent variables given $X_1, \ldots, X_n$, this yields, for any $k \in \{4, 6\}$,

$$\mathbb{E}_n\left[\Sigma_{in}^k\right] \leq C\mathbb{E}\left[\varepsilon^k\right]\left[\frac{\mathbb{1}\left(X_i \in \mathcal{X}_0\right)}{(nb_0^d)^2\widehat{g}_{in}^2}\sum_{j=1}^n K_0^2\left(\frac{X_j - X_i}{b_0}\right)\right]^{k/2} \leq \frac{C\mathbb{1}\left(X_i \in \mathcal{X}_0\right)\widetilde{g}_{in}^{k/2}}{(nb_0^d)^{(k/2)}\widehat{g}_{in}^k}.$$

Hence this bound and (A.16) give

$$\mathbb{E}_n\left[\mathbb{1}(X_i \in \mathcal{X}_0)\left(\widehat{m}_{in} - m(X_i)\right)^k\right] \leq C\left[\beta_{in}^k + \frac{\mathbb{1}\left(X_i \in \mathcal{X}_0\right)\widetilde{g}_{in}^{k/2}}{(nb_0^d)^{(k/2)}\widehat{g}_{in}^k}\right],$$

which proves (A.15), and then completes the proof of the lemma. □

Proof of Lemma 3.11

Since $K_0(\cdot)$ has a compact support under (A_7), there is a $C > 0$ such that $\|X_i - X_j\| \geq Cb_0$ implies that for any integer number k of $[1, n]$, $K_0((X_k - X_i)/b_0) = 0$ if $K_0((X_j - X_k)/b_0) \neq 0$. Let $D_j \subset [1, n]$ be such that an integer number k of $[1, n]$ is in D_j if and only if $K_0((X_j - X_k)/b_0) \neq 0$. Abbreviate $\mathbb{P}(\cdot|X_1, \ldots, X_n)$ into $\mathbb{P}_n$ and assume that $\|X_i - X_j\| \geq Cb_0$ so that D_i and D_j have an

empty intersection. Note also that taking C large enough ensures that i is not in D_j and j is not in D_i. It then follows, under (A_4) and since D_i and D_j only depend upon $X_1,\ldots,X_n$,

$$\begin{aligned}
&\mathbb{P}_n\Big((\widehat{m}_{in}-m(X_i),\varepsilon_i)\in A \text{ and } (\widehat{m}_{jn}-m(X_j),\varepsilon_j)\in B\Big)\\
&= \mathbb{P}_n\left(\left(\frac{\sum_{k\in D_i\setminus\{i\}}(m(X_k)-m(X_i)+\varepsilon_k)K_0((X_k-X_i)/b_0)}{\sum_{k\in D_i\setminus\{i\}}K_0((X_k-X_i)/b_0)},\varepsilon_i\right)\in A\right.\\
&\qquad\qquad \left.\text{and } \left(\frac{\sum_{\ell\in D_j\setminus\{j\}}(m(X_\ell)-m(X_j)+\varepsilon_\ell)K_0((X_\ell-X_j)/b_0)}{\sum_{\ell\in D_j\setminus\{j\}}K_0((X_\ell-X_j)/b_0)},\varepsilon_j\right)\in B\right)\\
&= \mathbb{P}_n\left(\left(\frac{\sum_{k\in D_i\setminus\{i\}}(m(X_k)-m(X_i)+\varepsilon_k)K_0((X_k-X_i)/b_0)}{\sum_{k\in D_i\setminus\{i\}}K_0((X_k-X_i)/b_0)},\varepsilon_i\right)\in A\right)\\
&\qquad\qquad \times\mathbb{P}_n\left(\left(\frac{\sum_{\ell\in D_j\setminus\{j\}}(m(X_\ell)-m(X_j)+\varepsilon_\ell)K_0((X_\ell-X_j)/b_0)}{\sum_{\ell\in D_j\setminus\{j\}}K_0((X_\ell-X_j)/b_0)},\varepsilon_j\right)\in B\right)\\
&= \mathbb{P}_n((\widehat{m}_{in}-m(X_i),\varepsilon_i)\in A)\times\mathbb{P}_n((\widehat{m}_{jn}-m(X_j),\varepsilon_j)\in B).
\end{aligned}$$

This gives the result of Lemma 3.11, since both $(\widehat{m}_{in}-m(X_i),\varepsilon_i)$ and $(\widehat{m}_{jn}-m(X_j),\varepsilon_j)$ are independent given $X_1,\ldots,X_n$. □

Proof of Lemma 3.12

Since $\widehat{m}_{in}-m(X_i)$ depends only upon $(X_1,\ldots,X_n,\varepsilon_k,k\neq i)$, we have

$$\sum_{i=1}^n \operatorname{Var}_n(\zeta_{in}) \le \sum_{i=1}^n \mathbb{E}_n\left[\zeta_{in}^2\right] = \sum_{i=1}^n \mathbb{E}_n\left[\mathbb{1}(X_i\in\mathcal{X}_0)(\widehat{m}_{in}-m(X_i))^4\,\mathbb{E}_{in}\left[K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)^2\right]\right],$$

with, using Lemma 3.7-(3.6.2),

$$\mathbb{E}_{in}\left[K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)^2\right] = \int K_1^{(2)}\left(\frac{e-\epsilon}{b_1}\right)^2 f(e)de \le Cb_1.$$

Therefore these bounds and Lemma 3.10 give

$$\begin{aligned}
\sum_{i=1}^n \operatorname{Var}_n(\zeta_{in}) &\le Cb_1\sum_{i=1}^n \mathbb{E}_n\Big[\mathbb{1}(X_i\in\mathcal{X}_0)(\widehat{m}_{in}-m(X_i))^4\Big]\\
&\le Cnb_1\sup_{1\le i\le n}\mathbb{E}_n\Big[\mathbb{1}(X_i\in\mathcal{X}_0)(\widehat{m}_{in}-m(X_i))^4\Big]\\
&\le O_{\mathbb{P}}(nb_1)\left(b_0^4+\frac{1}{nb_0^d}\right)^2.
\end{aligned}$$

which yields the desired result for the conditional variance.

We now prepare to compute the order of the conditional covariance. To that aim, observe that Lemma 3.11 gives

$$\sum_{i=1}^n\sum_{\substack{j=1\\j\neq i}}^n \operatorname{Cov}_n(\zeta_{in},\zeta_{jn}) = \sum_{i=1}^n\sum_{\substack{j=1\\j\neq i}}^n \mathbb{1}\Big(\|X_i-X_j\|<Cb_0\Big)\Big(\mathbb{E}_n[\zeta_{in}\zeta_{jn}]-\mathbb{E}_n[\zeta_{in}]\mathbb{E}_n[\zeta_{jn}]\Big).$$

The order of the term above is derived from the following equalities :

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathbb{1}\Big(\|X_i-X_j\|<Cb_0\Big)\mathbb{E}_n\left[\zeta_{in}\right]\mathbb{E}_n\left[\zeta_{jn}\right] = O_{\mathbb{P}}\left(n^2b_0^db_1^6\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^2, \quad (A.17)$$

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathbb{1}\Big(\|X_i-X_j\|<Cb_0\Big)\mathbb{E}_n\left[\zeta_{in}\zeta_{jn}\right] = O_{\mathbb{P}}\left(n^2b_0^db_1^{7/2}\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^2. \quad (A.18)$$

Indeed, since b_1 goes to 0 under (A_{10}), (A.17) and (A.18) yield that

$$\begin{aligned}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathrm{Cov}_n\left(\zeta_{in},\zeta_{jn}\right) &= O_{\mathbb{P}}\left[\left(n^2b_0^db_1^6\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^2+\left(n^2b_0^db_1^{7/2}\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^2\right]\\ &= O_{\mathbb{P}}\left(n^2b_0^db_1^{7/2}\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^2,\end{aligned}$$

which gives the result for the conditional covariance. Hence, it remains to prove (A.17) and (A.18). For (A.17), note that by (A_4) and Lemma 3.7-(3.6.2), we have

$$\begin{aligned}\left|\mathbb{E}_n\left[\zeta_{in}\right]\right| &= \left|\mathbb{E}_n\left[\mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left(\widehat{m}_{in}-m(X_i)\right)^2\mathbb{E}_{in}\left[K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\right]\right|\\ &\leq Cb_1^3\Bigg(\mathbb{E}_n\left[\mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left(\widehat{m}_{in}-m(X_i)\right)^4\right]\Bigg)^{1/2}.\end{aligned}$$

Hence from this bound and Lemma 3.10 we deduce

$$\begin{aligned}\sup_{1\leq i,j\leq n}\left|\mathbb{E}_n\left[\zeta_{in}\right]\mathbb{E}_n\left[\zeta_{jn}\right]\right| &\leq Cb_1^6\sup_{1\leq i\leq n}\mathbb{E}_n\left[\mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left(\widehat{m}_{in}-m(X_i)\right)^4\right]\\ &\leq O_{\mathbb{P}}\left(b_1^6\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^2.\end{aligned}$$

Therefore, since the Markov inequality gives

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathbb{1}\Big(\|X_i-X_j\|<Cb_0\Big)=O_{\mathbb{P}}(n^2b_0^d), \quad (A.19)$$

it then follows that

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathbb{1}\Big(\|X_i-X_j\|<Cb_0\Big)\mathbb{E}_n\left[\zeta_{in}\right]\mathbb{E}_n\left[\zeta_{jn}\right]=O_{\mathbb{P}}\left(n^2b_0^db_1^6\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^2,$$

which proves (A.17).

For (A.18), set $Z_{in}=\mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left(\widehat{m}_{in}-m(X_i)\right)^2$, and note that for $i\neq j$, we have

$$\mathbb{E}_n\left[\zeta_{in}\zeta_{jn}\right]=\mathbb{E}_n\left[Z_{in}K_1^{(2)}\left(\frac{\varepsilon_j-\epsilon}{b_1}\right)\mathbb{E}_{in}\left[Z_{jn}K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\right], \quad (A.20)$$

where

$$\begin{aligned}&\mathbb{E}_{in}\left[Z_{jn}K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\\ &= \beta_{jn}^2\mathbb{E}_{in}\left[K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]+2\beta_{jn}\mathbb{E}_{in}\left[\Sigma_{jn}K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]+\mathbb{E}_{in}\left[\Sigma_{jn}^2K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right].\end{aligned} \quad (A.21)$$

The first term of Equality (A.21) is treated by using Lemma 3.7-(3.6.2). This gives

$$\left|\beta_{jn}^2 \mathbb{E}_{in}\left[K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\right| \leq C b_1^3 \beta_{jn}^2. \tag{A.22}$$

Since under (A_4), the ε_j's are independent centered variables, and are independent of the X_j's, the second term in (A.21) gives

$$\begin{aligned}
\mathbb{E}_{in}\left[\Sigma_{jn} K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right] &= \frac{\mathbb{1}\left(X_j \in \mathcal{X}_0\right)}{n b_0^d \widehat{g}_{jn}} \sum_{k=1, k\neq j}^{n} K_0\left(\frac{X_k-X_j}{b_0}\right) \mathbb{E}_{in}\left[\varepsilon_k K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right] \\
&= \frac{\mathbb{1}\left(X_j \in \mathcal{X}_0\right)}{n b_0^d \widehat{g}_{jn}} K_0\left(\frac{X_i-X_j}{b_0}\right) \mathbb{E}_{in}\left[\varepsilon_i K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right].
\end{aligned}$$

Therefore, by (A_7) which ensures that K_0 is bounded, the equality above and Lemma 3.7-(3.6.2) yield that

$$\left|\beta_{jn} \mathbb{E}_{in}\left[\Sigma_{jn} K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\right| \leq C b_1^3 \left|\beta_{jn} \frac{\mathbb{1}\left(X_j \in \mathcal{X}_0\right)}{n b_0^d \widehat{g}_{jn}}\right|. \tag{A.23}$$

For the last term in (A.21), we have

$$\begin{aligned}
&\mathbb{E}_{in}\left[\Sigma_{jn}^2(x) K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right] \\
&= \frac{1}{(n b_0^d \widehat{g}_{jn})^2} \sum_{\substack{k=1 \\ k\neq j}}^{n} \sum_{\substack{\ell=1 \\ \ell\neq j}}^{n} K_0\left(\frac{X_k-X_j}{b_0}\right) K_0\left(\frac{X_\ell-X_j}{b_0}\right) \mathbb{E}_{in}\left[\varepsilon_k \varepsilon_\ell K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right] \\
&= \frac{1}{(n b_0^d \widehat{g}_{jn})^2} \sum_{k=1, k\neq j}^{n} K_0^2\left(\frac{X_k-X_j}{b_0}\right) \mathbb{E}_{in}\left[\varepsilon_k^2 K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right],
\end{aligned}$$

with, using Lemma 3.7-(3.6.2),

$$\begin{aligned}
&\left|\mathbb{E}_{in}\left[\varepsilon_k^2 K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\right| \\
&\leq \max\left\{\sup_{e\in\mathbb{R}}\left|\mathbb{E}_{in}\left[\varepsilon^2 K_1^{(2)}\left(\frac{\varepsilon-e}{b_1}\right)\right]\right|, \mathbb{E}[\varepsilon^2] \sup_{e\in\mathbb{R}}\left|\mathbb{E}_{in}\left[K_1^{(2)}\left(\frac{\varepsilon-e}{b_1}\right)\right]\right|\right\} \\
&\leq C b_1^3.
\end{aligned}$$

Therefore

$$\left|\mathbb{E}_{in}\left[\Sigma_{jn}^2 K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\right| \leq \frac{C b_1^3}{(n b_0^d \widehat{g}_{jn})^2} \sum_{k=1, k\neq j}^{n} K_0^2\left(\frac{X_k-X_j}{b_0}\right).$$

Substituting this bound, (A.23) and (A.22) in (A.21), we obtain

$$\left|\mathbb{E}_{in}\left[Z_{jn} K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\right| \leq C b_1^3 M_n,$$

where

$$M_n = \sup_{1\leq j\leq n}\left[\beta_{jn}^2 + \left|\beta_{jn} \frac{\mathbb{1}\left(X_j \in \mathcal{X}_0\right)}{n b_0^d \widehat{g}_{jn}}\right| + \frac{1}{(n b_0^d \widehat{g}_{jn})^2} \sum_{k=1, k\neq j}^{n} K_0^2\left(\frac{X_k-X_j}{b_0}\right)\right].$$

Hence from (A.20), the Cauchy-Schwarz inequality, Lemma 3.10 and Lemma 3.7-(3.6.2), we deduce

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} \mathbb{1}\Big(\|X_i - X_j\| < Cb_0\Big)\left|\mathbb{E}_n\left[\zeta_{in}\zeta_{jn}\right]\right| \\
&\leq \;\; CM_n b_1^3 \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} \mathbb{1}\Big(\|X_i - X_j\| < Cb_0\Big)\mathbb{E}_n\left|Z_{in}K_1^{(2)}\left(\frac{\varepsilon_j - \epsilon}{b_1}\right)\right| \\
&\leq \;\; CM_n b_1^3 \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} \mathbb{1}\Big(\|X_i - X_j\| < Cb_0\Big)\mathbb{E}_n^{1/2}\left[Z_{in}^2\right]\mathbb{E}_n^{1/2}\left[K_1^{(2)}\left(\frac{\varepsilon_j - \epsilon}{b_1}\right)^2\right] \\
&\leq \;\; M_n b_1^3 O_{\mathbb{P}}\left(b_0^4 + \frac{1}{nb_0^d}\right)(b_1)^{1/2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\Big(\mathbb{1}\left(\|X_i - X_j\| \leq Cb_0\right)\Big).
\end{aligned}$$

Moreover, (A.8) and Lemma 3.1 give, under (A_1), (A_7) and (A_9),

$$M_n = O_{\mathbb{P}}\left(b_0^4 + \frac{b_0^2}{nb_0^d} + \frac{1}{nb_0^d}\right) = O_{\mathbb{P}}\left(b_0^4 + \frac{1}{nb_0^d}\right).$$

Finally, substituting this order in the bound above, and using (A.19), we arrive at

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} \mathbb{1}\Big(\|X_i - X_j\| < Cb_0\Big)\mathbb{E}_n\left[\zeta_{in}\zeta_{jn}\right] = O_{\mathbb{P}}\left(n^2 b_0^d b_1^{7/2}\right)\left(b_0^4 + \frac{1}{nb_0^d}\right)^2.$$

This proves (A.18), and then completes the proof of the theorem. □

Chapitre 4

An integral nonparametric kernel estimator of the probability density function of regression errors

Abstract : This chapter is devoted to the nonparametric density estimation of the regression error using an integral method. The difference between the feasible estimator which uses the estimated regression function and the unfeasible one using the true regression function is investigated. An optimal choice of the first-step bandwidth used for estimating this regression function is proposed. We also study the asymptotic normality of the feasible integral kernel estimator and its rate-optimality.

4.1 Introduction

Consider a sample $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$ of independent and identically distributed (i.i.d) random variables, where Y is the univariate dependent variable and the covariate X is of dimension d. Let $m(\cdot)$ be the conditional expectation of Y given X and let ε be the related regression error term, so that the regression error model is

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, \ldots, n. \tag{4.1.1}$$

The aim of this chapter is to estimate the p.d.f of the regression error under the assumption that the covariate X and the regression error ε are independent. Indeed, under this assumption, we have

$$f(\epsilon) = f(\epsilon|x) = \varphi\left(m(x) + \epsilon|x\right). \tag{4.1.2}$$

Hence, the approach proposed here is based on a two-steps procedure, which, in a first step, uses (4.1.2) and writes $f(\epsilon)$ in the integral form

$$f(\epsilon) = \int \mathbb{1}\left(x \in \mathcal{X}\right) \varphi\left(\epsilon + m(x) \mid x\right) g(x) dx = \int \mathbb{1}\left(x \in \mathcal{X}\right) \varphi\left(x, \epsilon + m(x)\right) dx.$$

where $\mathcal{X}$ is the support of the p.d.f $g(\cdot)$ of X, and $\varphi(\cdot,\cdot)$ the joint density of (X,Y). This formula suggests to estimate $f(\epsilon)$, in a second-step, by

$$\widehat{f}_{2n}(\epsilon) = \int \mathbb{1}\left(x \in \mathcal{X}\right) \widehat{\varphi}_n\left(x, \epsilon + \widehat{m}_n(x)\right) dx,$$

where $\widehat{\varphi}$ and $\widehat{m}_n$ define respectively some nonparametric estimators of φ and m. As in Chapter 2, a challenging issue is first to evaluate the impact of the estimated regression function on the final estimator of $f(\cdot)$. Next, an optimal choice of the bandwidth used to estimate the residuals is proposed. Finally, we study the asymptotic normality of the estimator $\widehat{f}_{2n}(\epsilon)$ and its rate-optimality.

The rest of this chapter is organized as follows. Section 4.2 is devoted to presentation of ours estimators. Sections 4.3 and 4.4 group our assumptions and main results. The conclusion of this paper is given in Section 4.5, while the proofs of our results are gathered in section 4.6 and in two appendixes.

4.2 Presentation of the estimators

In what follows, the bandwidths b_0 and b_1 are associated with X and h with Y, and K_0, K_1 and K_2 represent some Kernels functions. Then for $(x,y) \in \mathbb{R}^d \times \mathbb{R}$, the nonparametric estimators of $\varphi(x,y)$ and $g(x)$ are respectively defined as

$$\begin{aligned}
\widehat{\varphi}_n\left(x,y\right) &= \frac{1}{nb_1^d h}\sum_{i=1}^n K_1\left(\frac{X_i - x}{b_1}\right) K_2\left(\frac{Y_i - y}{h}\right), \\
\widehat{g}_n\left(x\right) &= \frac{1}{nb_0^d}\sum_{i=1}^n K_0\left(\frac{X_i - x}{b_0}\right).
\end{aligned}$$

The estimation of the regression function $m(\cdot)$ is given by the Nadaraya-Watson estimator (1964)

$$\widehat{m}_n(x) = \frac{\sum_{j=1}^n Y_j K_0\left(\frac{X_j - x}{b_0}\right)}{\sum_{j=1}^n K_0\left(\frac{X_j - x}{b_0}\right)}. \tag{4.2.3}$$

Since $Y = m(X) + \varepsilon$, we have

$$\mathbb{P}\left(\varepsilon \leq \epsilon \mid X = x\right) = \mathbb{P}\left(Y \leq \epsilon + m(x) \mid X = x\right).$$

Then if f represents the probability density function of ε, and φ the joint density of (X,Y), it follows

$$f(\epsilon) = \int \mathbb{1}\left(x \in \mathcal{X}\right) \varphi\left(\epsilon + m(x) | x\right) g(x) dx = \int \mathbb{1}\left(x \in \mathcal{X}\right) \varphi\left(x, \epsilon + m(x)\right) dx, \tag{4.2.4}$$

where $\mathcal{X}$ is the support of the p.d.f g of the covariates. Therefore an estimator of $f(\epsilon)$ is the so-called "Two-steps estimator", defined as

$$\widehat{f}_{2n}(\epsilon) = \int \mathbb{1}\,(x \in \mathcal{X})\,\widehat{\varphi}_n\,(x, \epsilon + \widehat{m}_n(x))\,dx. \tag{4.2.5}$$

This estimator is a feasible estimator in the sense that it does not depend on any unknown quantity, as desirable in practice. This contrasts with the unfeasible ideal Kernel estimator

$$\widetilde{f}_{2n}(\epsilon) = \int \mathbb{1}\,(x \in \mathcal{X})\,\widehat{\varphi}_n\,(x, \epsilon + m(x))\,dx, \tag{4.2.6}$$

which depends in particular on the unknown regression function $m(\cdot)$. It is however intuitively clear that $\widehat{f}_{2n}(\epsilon)$ and $\widetilde{f}_{2n}(\epsilon)$ should be closed, as illustrated by the results of the next section.

4.3 Assumptions

($\mathbf{H_1}$) *The support $\mathcal{X}$ of X is a known compact subset of $\mathbb{R}^d$,*

($\mathbf{H_2}$) *the p.d.f. $g(\cdot)$ of the i.i.d. covariates X, X_i has continuous second order partial derivatives over $\mathcal{X}$. Moreover, there exists $\alpha > 0$ such that $g(x) > \alpha$ for all x in the support $\mathcal{X}$,*

($\mathbf{H_3}$) *the regression function $m(\cdot)$ has continuous second order partial derivatives over $\mathcal{X}$,*

($\mathbf{H_4}$) *the i.i.d. centered error regression terms $\varepsilon, \varepsilon_i$'s, have finite 6th moments, and are independent of the covariates X, X_i's,*

($\mathbf{H_5}$) *the probability density function f of ε has bounded continuous second order derivatives over $\mathbb{R}$, and satisfies, for $h_p(e) = e^p f(e)$, $\sup_{e\in\mathbb{R}} |h_p^{(k)}(e)| < \infty$, $p \in [0, 6]$, $k \in [0, 2]$,*

($\mathbf{H_6}$) *the density φ of (X, Y) has bounded continuous second order partial derivatives over $\mathbb{R}^d \times \mathbb{R}$,*

($\mathbf{H_7}$) *the Kernel functions K_0 and K_1 are symmetric, continuous over $\mathbb{R}^d$ with support in $[-1/2, 1/2]^d$ and $\int K_0(z)dz = 1$, $\int K_1(z)dz = 1$,*

($\mathbf{H_8}$) *the Kernel function K_2 has a compact support, is three times continuously differentiable over $\mathbb{R}$, and satisfies $\int K_2(v)dv = 1$, $\int vK_2(v)dv = 0$ and $\int |v^p K_2^{(\ell)}(v)|dv < \infty$ for p, ℓ in $[0, 3]$,*

($\mathbf{H_9}$) *the bandwidth b_0 decreases to 0 and satisfies $\ln(1/b_0)/\ln(\ln n) \to \infty$ and $b_0^d/(nb_0^{2d})^p = O(b_0^{2p})$, $p \in [0, 6]$, when $n \to \infty$,*

($\mathbf{H_{10}}$) *the bandwidths b_1 and h decrease to 0 and are such that $nb_1^{2d} \to \infty$ and $n^{(d+8)}h^{7(d+4)} \to \infty$ when $n \to \infty$.*

Assumptions (H_2), (H_3), (H_5) and (H_6) impose that all the functions to be estimated nonparametrically have two bounded derivatives. Consequently the conditions $\int vK_j(v)dv = 0$, $j = 0, 1, 2$, as assumed in (H_7) and (H_8), represent standard conditions ensuring that the bias of the resulting

nonparametric estimators (4.2.3) and (4.2.6) are respectively of order b_0^2 and $b_0^2 + h^2$. Assumption (H_4) states independence between the regression error terms and the covariates, which is the main condition for (4.1.2) to hold. The differentiability of K_2 imposed in (H_8) is more specific to our two-steps estimation method. Assumption (H_8) is used to expand the two-steps Kernel estimator $\widehat{f}_{2n}$ in (4.2.5) around the unfeasible one $\widetilde{f}_{2n}$ from (4.2.6), using the derivatives of K_2 up to third order and the differences $\widehat{m}_{in}(x) - m(x)$, $i \in [1, n]$, where $\widehat{m}_{in}(x)$ is a leave-one out version of the Kernel regression estimator (4.2.3),

$$\widehat{m}_{in}(x) = \frac{\sum_{\substack{j=1 \\ j \neq i}}^{n} Y_j K_0\left(\frac{X_j - x}{b_0}\right)}{\sum_{\substack{j=1 \\ j \neq i}}^{n} K_0\left(\frac{X_j - x}{b_0}\right)}. \tag{4.3.7}$$

Assumption (H_9) is a standard condition to obtain uniform convergence of the regression estimator $\widehat{m}_n$ in (4.2.3) (see for instance Einmahl and Mason, 2005), and also gives a similar consistency result for the leave-one-out estimator $\widehat{m}_{in}$. Assumption (H_{10}) is needed in the study of the difference between the feasible estimator $\widehat{f}_{2n}$ and the unfeasible estimator $\widetilde{f}_{2n}$.

4.4 Main results

Our first main result establishes the order of the difference $\widehat{f}_{2n}(\epsilon) - f(\epsilon)$. This is given in the following subsection. Next, we shall give the optimal bandwidths needed to estimate $f(\epsilon)$. We conclude this section by proposing an asymptotic normality of the estimator $\widehat{f}_{2n}(\epsilon)$.

4.4.1 Pointwise weak consistency

In this subsection we deal the order of the difference $\widehat{f}_{2n}(\epsilon) - f(\epsilon)$. We show that for n large enough, the estimator $\widehat{f}_{2n}(\epsilon)$ is very close to the theoretical density $f(\epsilon)$, as illustrated by the following result.

Theorème 4.1. *Suppose that Assumptions $(H_1) - (H_{10})$ hold. Then for n large enough, we have*

$$\widehat{f}_{2n}(\epsilon) - f(\epsilon) = O_{\mathbb{P}}\Big(AMSE(b_1, h) + RT_n(b_0, b_1, h)\Big)^{1/2},$$

where

$$AMSE(b_1, h) = \mathbb{E}_n\left[\left(\widetilde{f}_{2n}(\epsilon) - f(\epsilon)\right)^2\right] = O_{\mathbb{P}}\left(b_1^4 + h^4 + \frac{1}{nb_1}\right),$$

and

$$\begin{aligned} RT_n(b_0, b_1, h) \quad = \quad & b_0^4 + \left(b_0^d \vee b_1^d\right)\left[\frac{1}{nb_1^d h^3}\left(b_0^4 + \frac{1}{nb_0^d}\right) + \frac{1}{nb_0^d}\right] \\ & + \left(b_0^d \vee b_1^d\right)\left[\frac{1}{nb_1^d h^5}\left(b_0^4 + \frac{1}{nb_0^d}\right)^2 + \frac{1}{n^2 b_0^{2d} h^3}\right] \\ & + \frac{1}{h^2}\left(b_0^4 + \frac{1}{nb_0^d}\right)^3 + \frac{b_0^d \vee b_1^d}{h^7}\left(b_0^4 + \frac{1}{nb_0^d}\right)^3. \end{aligned}$$

The result of Theorem 4.1 is based on the evaluation of the difference between $\widehat{f}_{2n}(\epsilon)$ and $\widetilde{f}_{2n}(\epsilon)$. This evaluation gives an indication about the impact of the estimation of $m(\cdot)$ on the nonparametric estimation of the regression error density.

4.4.2 Optimal first-step and second-step bandwidths for the pointwise weak consistency

Our next result deals with the choice of the optimal bandwidth b_0 used in the nonparametric estimation of the p.d.f of the regression error term. We have the following theorem.

Theorem 4.2. *Suppose that Assumptions $(H_1)-(H_{10})$ are satisfied, and assume $b_0 = b_1$. Define*

$$b_0^* = b_0^*(h) = \arg\min_{b_0} RT_n(b_0, b_0, h),$$

where the minimization is performed over bandwidth b_0 fulfilling (H_9). Then the optimal bandwidth b_0^* satisfies

$$b_0^* \asymp \max\left\{\left(\frac{1}{n^2h^3}\right)^{\frac{1}{d+4}}, \left(\frac{1}{n^3h^7}\right)^{\frac{1}{2d+4}}\right\},$$

and we have

$$RT_n(b_0^*, b_0^*, h) \asymp \frac{1}{n} + \max\left\{\left(\frac{1}{n^2h^3}\right)^{\frac{4}{d+4}}, \left(\frac{1}{n^3h^7}\right)^{\frac{4}{2d+4}}\right\}.$$

The next theorem gives the conditions for which the estimator $\widehat{f}_{2n}(\epsilon)$ reaches the optimal rate $n^{-2/5}$ when b_0 takes the value b_0^*. We prove that for $d \leq 2$, the bandwidth that minimizes the term $AMSE(b_0^*, h) + RT_n(b_0^*, b_0^*, h)$ has the same order as $n^{-1/5}$, leading to the optimal order $n^{-2/5}$ for the term $(AMSE(b_0^*, h) + RT_n(b_0^*, b_0^*, h))^{1/2}$.

Theorem 4.3. *Assume that $(H_1)-(H_{10})$ hold and set*

$$h^* = \arg\min_h \Big(AMSE(b_0^*, h) + RT_n(b_0^*, b_0^*, h)\Big),$$

where $b_0^* = b_0^*(h)$ is defined as in Theorem 4.2. Then

1. *For $d \leq 2$, the optimal bandwidth h^* satisfies*

$$h^* \asymp \left(\frac{1}{n}\right)^{\frac{1}{5}},$$

and we have

$$\Big(AMSE(b_0^*, h^*) + RT_n\left(b_0^*, h^*, h^*\right)\Big)^{\frac{1}{2}} \asymp \left(\frac{1}{n}\right)^{\frac{2}{5}}.$$

2. *For $d \geq 3$, h^* satisfies*

$$h^* \asymp \left(\frac{1}{n}\right)^{\frac{3}{2d+11}},$$

and we have

$$\Big(AMSE(b_0^*, h^*) + RT_n(b_0^*, b_0^*, h^*)\Big)^{\frac{1}{2}} \asymp \left(\frac{1}{n}\right)^{\frac{6}{2d+11}}.$$

Theorem 4.3 follows from Theorem 4.2, which reveals that for b_1 proportional to $n^{-1/5}$, the bandwidth b_0^* has the same order as

$$\max\left\{\left(\frac{1}{n}\right)^{\frac{7}{5(d+4)}}, \left(\frac{1}{n}\right)^{\frac{8}{5(2d+4)}}\right\} = \left(\frac{1}{n}\right)^{\frac{8}{5(2d+4)}}.$$

For $d \leq 2$, this order of b_0^* is less than the one of the optimal bandwidth $\widehat{b}_0$ obtained for pointwise or mean square estimation of $m(\cdot)$ using a nonparametric Kernel estimator. In fact, as seen in Chapter 3, the optimal bandwidth $\widehat{b}_0$ for estimating $m(\cdot)$ is obtained by minimizing the order of the risk function

$$r_n(b_0) = \mathbb{E}\left[\int \mathbb{1}\left(x \in \mathcal{X}\right)\left(\widehat{m}_n(x) - m(x)\right)^2 \widehat{g}_n^2(x) w(x) dx\right],$$

which has the same order as $b_0^4 + \left(1/(nb_0^d)\right)$, leading to the optimal bandwidth $\widehat{b}_0 = n^{-1/(d+4)}$. For d=1, the optimal order of b_0^* is $n^{-(1/5)\times(4/3)}$ which goes to 0 slightly faster than $n^{-1/5}$, the optimal order of the bandwidth for the mean square nonparametric estimation of $m(\cdot)$. For $d = 2$, the optimal order of b_0^* is $n^{-1/5}$. Again this order goes to 0 faster than the order $n^{-1/6}$ of the optimal bandwidth for the nonparametric estimation of the regression function with two covariates. But for $d \geq 3$, we note that the order of b_0^* goes to 0 slowly than $\widehat{b}_0$. Hence these sitauations reveal that the optimal $\widehat{m}_n(\cdot)$ for estimating $f(\cdot)$ should have a lower bias and a higher variance than the optimal Kernel regression estimator of $m(\cdot)$. This situation is the same as the one noticed in Wang, Cai, Brown and Levine (2008) for the estimation of the conditional variance function in a heteroscedastic regression model. However these authors do not investigate the order of the optimal bandwidth to be used for estimating the regression function in their heteroscedastic setup. Hence, as in Chapter 3, we conclude that an estimator of $m(\cdot)$ with smaller bias should be preferred in our framework, compared to the case where the regression function $m(\cdot)$ is the parameter of interest.

4.4.3 Asymptotic normality

The aim of this subsection is to propose an asymptotic normality of the estimator $\widehat{f}_{2n}(\epsilon)$. We have the following result.

Theorem 4.4. *Suppose that $b_0 = b_1$ and assume*

$$(\mathbf{H_{11}}): \quad nb_0^{d+4} = O(1), \quad nb_0^4 h = o(1), \quad nb_0^d h^3 \to \infty,$$

when $n \to \infty$. Then under $(H_1) - (H_{10})$, we have

$$\sqrt{nh}\left(\widehat{f}_{2n}(\epsilon) - \overline{f}_{2n}(\epsilon)\right) \xrightarrow{d} \mathcal{N}\left(0, f(\epsilon)\int K_2^2(v)dv\right),$$

where

$$\begin{aligned}\overline{f}_{2n}(\epsilon) &= f(\epsilon) + \frac{b_0^2}{2}\int \mathbb{1}\left(x \in \mathcal{X}\right)\frac{\partial^2 \varphi(x, \epsilon + m(x))}{\partial^2 x} dx \int z K_1(z) z^\top dz \\ &\quad + \frac{h^2}{2}\int \mathbb{1}\left(x \in \mathcal{X}\right)\frac{\partial^2 \varphi(x, \epsilon + m(x))}{\partial^2 y} dx \int v^2 K_2(v) dv + o\left(b_0^2 + h^2\right).\end{aligned}$$

As seen in the comments of Theorem 3.4 in Chapter 3, we can check that for $d = 1$, $h = h^*$ and $b_1 = b_0 = b_0^*$, the conditions of Assumption $(\mathbf{H_{11}})$ are realizable with the bandwidths b_0^* and h^*. But with these bandwidths, the last constraint of $(\mathbf{H_{11}})$ is not satisfied for $d = 2$, since for $b_0 = b_0^*$ and $h = h^*$, $nb_0^d h^3$ is bounded when n goes to infinity.

4.5 Conclusion

In this chapter, we investigated the nonparametric Kernel estimation of the p.d.f of the regression error using an integral method. The difference between the feasible estimator which uses the estimated regression function and the unfeasible one using the theoretical regression function is studied. An optimal choice of the first-step bandwidth used to estimate the regression function is also established. Again, an asymptotic normality of the feasible Kernel estimator and its rate-optimality are proposed. As in Chapter 2, the contributions of the present chapter is the analysis of the influence of the estimated regression function on the regression errors p.d.f. Kernel estimator.

The strategy used here strategy is to use an approach based on a two-steps procedure which, in a first step, integrates a conditional p.d.f as given in (4.2.4). In a second step, we build the Kernel estimator of $f(\epsilon)$ by estimating nonparametrically the unknown functions in the integral terms of (4.2.4). If this strategy can avoid the curse of dimensionality, a main aspect of our setup is to evaluate the impact of the estimation of $m(\cdot)$ on the final integral Kernel estimator of $f(\cdot)$ in the first nonparametric step, and to determine the optimal choice of the first-step bandwidth b_0. For a such choice of b_0, our results suggests that the optimal bandwidth to be used should be smaller than the optimal bandwidth for the mean square estimation of $m(\cdot)$. This mean that the best choice for b_0 is the one such that the estimator $\widehat{m}_n(\cdot)$ of the regression has a lower bias and a higher variance than the optimal Kernel regression of the estimation setup. With this choice of b_0, we show that for $d \leq 2$, the estimator $\widehat{f}_{2n}(\epsilon)$ of $f(\epsilon)$ can reach the optimal rate $n^{-2/5}$, which corresponds exactly to the rate reached for the Kernel density estimator of an univariate variable. This reveals that for $d \leq 2$, the integral Kernel estimator $\widehat{f}_{2n}(\epsilon)$ is not affected by the curse of dimensionality, since there is not a negative influence caused by the estimation of the optimal first-step bandwidth b_0^*.

4.6 Proofs section

Proof of Theorem 4.1

The proof is a consequence of the two followings lemmas.

Lemma 4.1. *Under $(H_1) - (H_{10})$, we have, when n goes to infinity,*

$$\begin{aligned}\widehat{f}_{2n}(\epsilon) - \widetilde{f}_{2n}(\epsilon) &= O_{\mathbb{P}}\left[b_0^4 + \left(b_0^d \vee b_1^d\right)\left(\frac{1}{nb_1^d h^3}\left(b_0^4 + \frac{1}{nb_0^d}\right) + \frac{1}{nb_0^d}\right)\right]^{1/2} \\ &+ O_{\mathbb{P}}\left[\left(b_0^d \vee b_1^d\right)\left(\frac{1}{nb_1^d h^5}\left(b_0^4 + \frac{1}{nb_0^d}\right)^2 + \frac{1}{n^2 b_0^{2d} h^3}\right)\right]^{1/2} \\ &+ O_{\mathbb{P}}\left[\frac{1}{h^2}\left(b_0^4 + \frac{1}{nb_0^d}\right)^3 + \frac{b_0^d \vee b_1^d}{h^7}\left(b_0^4 + \frac{1}{nb_0^d}\right)^3\right]^{1/2}.\end{aligned}$$

Lemma 4.2. *If $(H_1) - (H_{10})$ hold, then*

$$\widetilde{f}_{2n}(\epsilon) - f(\epsilon) = O_{\mathbb{P}}\left(b_1^4 + h^4 + \frac{1}{nh}\right)^{1/2}.$$

Let now turn to the proof of Theorem 4.1. Using Lemmas 4.2 and 4.1, we have

$$\begin{aligned}&\widehat{f}_{2n}(\epsilon) - f(\epsilon) = \left(\widetilde{f}_{2n}(\epsilon) - f(\epsilon)\right) + \widehat{f}_{2n}(\epsilon) - \widetilde{f}_{2n}(\epsilon) \\ &= O_{\mathbb{P}}\left(b_1^4 + h^4 + \frac{1}{nh}\right)^{1/2} + O_{\mathbb{P}}\left[b_0^4 + \left(b_0^d \vee b_1^d\right)\left(\frac{1}{nb_1^d h^3}\left(b_0^4 + \frac{1}{nb_0^d}\right) + \frac{1}{nb_0^d}\right)\right]^{1/2} \\ &+ O_{\mathbb{P}}\left[\left(b_0^d \vee b_1^d\right)\left(\frac{1}{nb_1^d h^5}\left(b_0^4 + \frac{1}{nb_0^d}\right)^2 + \frac{1}{n^2 b_0^{2d} h^3}\right) + \frac{1}{h^2}\left(b_0^4 + \frac{1}{nb_0^d}\right)^3 + \frac{b_0^d \vee b_1^d}{h^7}\left(b_0^4 + \frac{1}{nb_0^d}\right)^3\right]^{1/2},\end{aligned}$$

which yields the result of the Theorem. □

We now prove Lemmas 4.1 and 4.2.

Proof of Lemma 4.1

Let us introduce additional notations. Let $\widehat{m}_{in}(x)$ be as in (3.2.4) and define

$$\begin{aligned}S_n(x) &= \frac{1}{nb_1^d h^2}\sum_{i=1}^n \left(\widehat{m}_{in}(x) - m(x)\right) K_1\left(\frac{X_i - x}{b_1}\right) K_2^{(1)}\left(\frac{Y_i - \epsilon - m(x)}{h}\right), \\ T_n(x) &= \frac{1}{nb_1^d h^3}\sum_{i=1}^n \left(\widehat{m}_{in}(x) - m(x)\right)^2 K_1\left(\frac{X_i - x}{b_1}\right) K_2^{(2)}\left(\frac{Y_i - \epsilon - m(x)}{h}\right).\end{aligned}$$

The proof of Lemma 4.1 is based on the following results.

Lemma 4.3. *Define*

$$S_n = \int \mathbb{1}\left(x \in \mathcal{X}\right) S_n(x)dx, \quad T_n = \int \mathbb{1}\left(x \in \mathcal{X}\right) T_n(x)dx.$$

Then under $(H_1)-(H_{10})$, we have

$$
\begin{aligned}
S_n &= O_{\mathbb{P}}\left[b_0^4+\left(b_0^d\vee b_1^d\right)\left(\frac{1}{nb_1^dh^3}\left(b_0^4+\frac{1}{nb_0^d}\right)+\frac{1}{nb_0^d}\right)\right]^{1/2},\\
T_n &= O_{\mathbb{P}}\left[\left(b_0^4+\frac{1}{nb_0^d}\right)^2+\left(b_0^d\vee b_1^d\right)\left(\frac{1}{nb_1^dh^5}\left(b_0^4+\frac{1}{nb_0^d}\right)^2+\frac{b_0^4}{nb_0^d}+\frac{1}{n^2b_0^{2d}h^3}\right)\right]^{1/2}.
\end{aligned}
$$

Lemma 4.4. *Define*

$$
R_n(x)=\frac{1}{nb_1^dh^4}\sum_{i=1}^n\left(\widehat{m}_{in}(x)-m(x)\right)^3K_1\left(\frac{X_i-x}{b_1}\right)\int_0^1(1-u)^2K_2^{(3)}\left(\frac{Y_i-\theta_{in}(x,u)}{h}\right)du,
$$

where $\theta_{in}(x,u)=\epsilon-m(x)-u\left(\widehat{m}_{in}(x)-m(x)\right)$, and set

$$
R_n=\int \mathbb{1}\left(x\in\mathcal{X}\right)R_n(x)dx.
$$

If $(H_1)-(H_{10})$ hold, then

$$
R_n=O_{\mathbb{P}}\left[\frac{1}{h^2}\left(b_0^4+\frac{1}{nb_0^d}\right)^3+\frac{b_0^d\vee b_1^d}{h^7}\left(b_0^4+\frac{1}{nb_0^d}\right)^3\right]^{1/2}.
$$

Lemma 4.5. *Set*

$$
P_n(x)=\frac{1}{nb_1^dh^2}\sum_{i=1}^n\left(\widehat{m}_n(x)-\widehat{m}_{in}(x)\right)K_1\left(\frac{X_i-x}{b_1}\right)\int_0^1K_2^{(1)}\left(\frac{Y_i-\widehat{\theta}_{in}(x,t)}{h}\right)dt,
$$

where $\widehat{\theta}_{in}(x,t)=\epsilon+\widehat{m}_{in}(x)+t\left(\widehat{m}_n(x)-\widehat{m}_{in}(x)\right)$, and define

$$
P_n=\int \mathbb{1}\left(x\in\mathcal{X}\right)P_n(x)dx.
$$

Then under $(H_1)-(H_{10})$, we have

$$
P_n=O_{\mathbb{P}}\left(\frac{1}{n^2b_0^{2d}}+\frac{b_0^d\vee b_1^d}{n^2b_0^{2d}h^3}\right)^{1/2}.
$$

The proofs of these Lemmas are stated in Appendix B.

Let us now return to the proof of Lemma 4.1. Observe that

$$\begin{aligned}
&\widehat{\varphi}_n\left(x,\epsilon+\widehat{m}_n(x)\right)-\widehat{\varphi}_n\left(x,\epsilon+m(x)\right)\\
&\quad = \frac{1}{nb_1^dh}\sum_{i=1}^{n}K_1\left(\frac{X_i-x}{b_1}\right)\left[K_2\left(\frac{Y_i-\epsilon-\widehat{m}_n(x)}{h}\right)-K_2\left(\frac{Y_i-\epsilon-m(x)}{h}\right)\right], \qquad (4.6.1)
\end{aligned}$$

where

$$\begin{aligned}
&K_2\left(\frac{Y_i-\epsilon-\widehat{m}_n(x)}{h}\right)-K_2\left(\frac{Y_i-\epsilon-m(x)}{h}\right)\\
&\quad = K_2\left(\frac{Y_i-\epsilon-\widehat{m}_{in}(x)}{h}\right)-K_2\left(\frac{Y_i-\epsilon-m(x)}{h}\right)\\
&\qquad +\left[K_2\left(\frac{Y_i-\epsilon-\widehat{m}_n(x)}{h}\right)-K_2\left(\frac{Y_i-\epsilon-\widehat{m}_{in}(x)}{h}\right)\right]. \qquad (4.6.2)
\end{aligned}$$

Since K_2 is three times continuously differentiable under (H_8), the Taylor's theorem with the integral remainder gives

$$\begin{aligned}
&K_2\left(\frac{Y_i-\epsilon-\widehat{m}_{in}(x)}{h}\right)-K_2\left(\frac{Y_i-\epsilon-m(x)}{h}\right)\\
&= -\frac{1}{h}\left(\widehat{m}_{in}(x)-m(x)\right)K_2^{(1)}\left(\frac{Y_i-\epsilon-m(x)}{h}\right)\\
&\quad +\frac{1}{2h^2}\left(\widehat{m}_{in}(x)-m(x)\right)^2K_2^{(2)}\left(\frac{Y_i-\epsilon-m(x)}{h}\right)\\
&\quad -\frac{1}{2h^3}\left(\widehat{m}_{in}(x)-m(x)\right)^3\int_0^1(1-u)^2K_2^{(3)}\left(\frac{Y_i-\epsilon-m(x)-u\left(\widehat{m}_{in}(x)-m(x)\right)}{h}\right)du. \qquad (4.6.3)
\end{aligned}$$

Again, under (H_8), we have

$$\begin{aligned}
&K_2\left(\frac{Y_i-\epsilon-\widehat{m}_n(x)}{h}\right)-K_2\left(\frac{Y_i-\epsilon-\widehat{m}_{in}(x)}{h}\right)\\
&= -\frac{1}{h}\left(\widehat{m}_n(x)-\widehat{m}_{in}(x)\right)\int_0^1K_2^{(1)}\left(\frac{Y_i-\epsilon-\widehat{m}_{in}(x)-t\left(\widehat{m}_n(x)-\widehat{m}_{in}(x)\right)}{h}\right)dt.
\end{aligned}$$

Hence defining $S_n(x)$, $T_n(x)$, $R_n(x)$ and $P_n(x)$ respectively as in Lemmas 4.3, 4.4 and 4.5, the equality above, (4.6.3), (4.6.2) and (4.6.1) give

$$\widehat{\varphi}_n\left(x,\epsilon+\widehat{m}_n(x)\right)-\widehat{\varphi}_n\left(x,\epsilon+m(x)\right)=-S_n(x)+\frac{T_n(x)}{2}-\frac{R_n(x)}{2}-P_n(x),$$

so that

$$\begin{aligned}
\widehat{f}_{2n}(\epsilon)-\widetilde{f}_{2n}(\epsilon) &= -S_n+\frac{T_n}{2}-\frac{R_n}{2}-P_n\\
&= O_{\mathbb{P}}\left[b_0^4+\left(b_0^d\vee b_1^d\right)\left(\frac{1}{nb_1^dh^3}\left(b_0^4+\frac{1}{nb_0^d}\right)+\frac{1}{nb_0^d}\right)\right]^{1/2}\\
&\quad +O_{\mathbb{P}}\left[\left(b_0^4+\frac{1}{nb_0^d}\right)^2+\left(b_0^d\vee b_1^d\right)\left(\frac{1}{nb_1^dh^5}\left(b_0^4+\frac{1}{nb_0^d}\right)^2+\frac{b_0^4}{nb_0^d}+\frac{1}{n^2b_0^{2d}h^3}\right)\right]^{1/2}\\
&\quad +O_{\mathbb{P}}\left[\frac{1}{h^2}\left(b_0^4+\frac{1}{nb_0^d}\right)^3+\frac{b_0^d\vee b_1^d}{h^7}\left(b_0^4+\frac{1}{nb_0^d}\right)^3+\frac{1}{n^2b_0^{2d}}+\frac{b_0^d\vee b_1^d}{n^2b_0^{2d}h^3}\right]^{1/2}.
\end{aligned}$$

Moreover, since under (H_9) b_0 goes to 0 and that $b_0^d/(n^p b_0^{2dp}) = O(b_0^{2p})$, this gives for $p = 1$,

$$\frac{b_0^4}{nb_0^d} = O\left(\frac{1}{nb_0^d}\right), \quad \frac{1}{n^2 b_0^{2d}} = O\left(b_0^4\right), \quad \left(b_0^4 + \frac{1}{nb_0^d}\right)^2 = O\left(b_0^4\right).$$

Hence it follows that

$$\begin{aligned}
\widehat{f}_{2n}(\epsilon) - \widetilde{f}_{2n}(\epsilon) &= O_{\mathbb{P}}\left[b_0^4 + \left(b_0^d \vee b_1^d\right)\left(\frac{1}{nb_1^d h^3}\left(b_0^4 + \frac{1}{nb_0^d}\right) + \frac{1}{nb_0^d}\right)\right]^{1/2} \\
&\quad + O_{\mathbb{P}}\left[\left(b_0^d \vee b_1^d\right)\left(\frac{1}{nb_1^d h^5}\left(b_0^4 + \frac{1}{nb_0^d}\right)^2 + \frac{1}{n^2 b_0^{2d} h^3}\right)\right]^{1/2} \\
&\quad + O_{\mathbb{P}}\left[\frac{1}{h^2}\left(b_0^4 + \frac{1}{nb_0^d}\right)^3 + \frac{b_0^d \vee b_1^d}{h^7}\left(b_0^4 + \frac{1}{nb_0^d}\right)^3\right]^{1/2},
\end{aligned}$$

which ends the proof of the Lemma. □

Proof of Lemma 4.2

Observe that

$$\widetilde{f}_{2n}(\epsilon) - f(\epsilon) = \left(\widetilde{f}_{2n}(\epsilon) - \mathbb{E}\widetilde{f}_{2n}(\epsilon)\right) + \left(\mathbb{E}\widetilde{f}_{2n}(\epsilon) - f(\epsilon)\right). \tag{4.6.4}$$

For the first term in (4.6.4), the independence of the (X_i, Y_i)'s gives

$$\begin{aligned}
&\mathbb{E}\left[\left(\widetilde{f}_{2n}(\epsilon) - \mathbb{E}\widetilde{f}_{2n}(\epsilon)\right)^2\right] = \mathrm{Var}\left(\widetilde{f}_{2n}(\epsilon)\right) \\
&= \mathrm{Var}\left[\frac{1}{nb_1^d h}\sum_{i=1}^n \int \mathbb{1}\left(x \in \mathcal{X}\right) K_1\left(\frac{X_i - x}{b_1}\right) K_2\left(\frac{Y_i - \epsilon - m(x)}{h}\right) dx\right] \\
&= \frac{1}{(nb_1^d h)^2}\sum_{i=1}^n \mathrm{Var}\left[\int \mathbb{1}\left(x \in \mathcal{X}\right) K_1\left(\frac{X_i - x}{b_1}\right) K_2\left(\frac{Y_i - \epsilon - m(x)}{h}\right) dx\right] \\
&\le \frac{1}{(nb_1^d h)^2}\sum_{i=1}^n \mathbb{E}\left[\int \mathbb{1}\left(x \in \mathcal{X}\right) K_1\left(\frac{X_i - x}{b_1}\right) K_2\left(\frac{Y_i - \epsilon - m(x)}{h}\right) dx\right]^2.
\end{aligned}$$

Moreover, note that by (H_1), (H_3) and $(H_7) - (H_8)$, the changes of variables $x = x_1 + hz_1$, $y_1 = \epsilon + m(x_1 + b_1 z_1) + hv_1$ and the Cauchy-Schwarz inequality give, since $\varphi(\cdot,\cdot)$ is bounded under Assumption (H_6),

$$\begin{aligned}
&\sum_{i=1}^n \mathbb{E}\left[\int \mathbb{1}\left(x \in \mathcal{X}\right) K_1\left(\frac{X_i - x}{b_1}\right) K_2\left(\frac{Y_i - \epsilon - m(x)}{h}\right) dx\right]^2 \\
&= n\int_{\mathbb{R}^d} dx_1 \int_{\mathbb{R}}\left[\int \mathbb{1}\left(x \in \mathcal{X}\right) K_1\left(\frac{x_1 - x}{b_1}\right) K_2\left(\frac{y_1 - \epsilon - m(x)}{h}\right) dx\right]^2 \varphi\left(x_1, y_1\right) dy_1 \\
&= n\int_{\mathbb{R}^d} dx_1 \int_{\mathbb{R}}\left[b_1^d \int \mathbb{1}\left(x_1 + b_1 z_1 \in \mathcal{X}\right) K_1(z_1) K_2\left(\frac{y_1 - \epsilon - m(x_1 + b_1 z_1)}{h}\right) dz_1\right]^2 \varphi\left(x_1, y_1\right) dy_1 \\
&\le Cnb_1^{2d} h \int_{\mathbb{R}^d} dz_1 K_1^2(z_1) \int \mathbb{1}\left(x_1 + b_1 z_1 \in \mathcal{X}\right) dx_1 \int_{\mathbb{R}} K_2^2(v_1) dv_1.
\end{aligned}$$

Hence from the two bounds above and the Tchebychev inequality, we deduce

$$\widetilde{f}_{2n}(\epsilon) - \mathbb{E}\widetilde{f}_{2n}(\epsilon) = O_{\mathbb{P}}\left(\frac{1}{nh}\right)^{1/2}. \tag{4.6.5}$$

We now compute the order of the second term in (4.6.4). Observe that

$$\begin{aligned}
\mathbb{E}\widetilde{f}_{2n}(\epsilon) &= \mathbb{E}\left[\frac{1}{nb_1^dh}\sum_{i=1}^{n}\int \mathbb{1}\left(x\in\mathcal{X}\right)K_1\left(\frac{X_i-x}{b_1}\right)K_2\left(\frac{Y_i-\epsilon-m(x)}{h}\right)dx\right] \\
&= \frac{n}{nb_1^dh}\int \mathbb{1}\left(x\in\mathcal{X}\right)\mathbb{E}\left[K_1\left(\frac{X_1-x}{b_1}\right)K_2\left(\frac{Y_1-\epsilon-m(x)}{h}\right)\right]dx \\
&= \int \mathbb{1}\left(x\in\mathcal{X}\right)\left[\int_{\mathbb{R}^d}dz_1\int_{\mathbb{R}}K_1(z_1)K_2(v_1)\varphi\left(x+b_1z_1,\epsilon+m(x)+hv_1\right)dv_1\right]dx. \quad (4.6.6)
\end{aligned}$$

By (H_6), a second-order Taylor expansion yields, for z_1 and v_1 in the supports of K_1 and K_2, and h and b_1 small enough,

$$\begin{aligned}
\varphi\left(x+b_1z_1,\epsilon+m(x)+hv_1\right) &= \varphi(x,\epsilon+m(x))+b_1\frac{\partial\varphi(x,\epsilon+m(x))}{\partial x}z_1^\top+h\frac{\partial\varphi(x,\epsilon+m(x))}{\partial y}v_1 \\
&\quad+\frac{b_1^2}{2}z_1\frac{\partial^2\varphi(x+\theta b_1z_1,\epsilon+m(x)+\theta b_1v_1)}{\partial^2x}z_1^\top \\
&\quad+b_1hv_1\frac{\partial^2\varphi(x+\theta b_1z_1,\epsilon+m(x)+\theta b_1v_1)}{\partial x\partial y}z_1^\top \\
&\quad+\frac{h^2}{2}\frac{\partial^2\varphi(x+\theta b_1z_1,\epsilon+m(x)+\theta b_1v_1)}{\partial^2y}v_1^2,
\end{aligned}$$

for some $\theta=\theta(x,\epsilon,b_1z_1,hv_1)$ in $[0,1]$. This gives, since $\int K_1(z)dz=\int K_2(v)dv=1$, $\int zK_1(z)dz$ and that $\int vK_2(v)dv$ vanishes under $(H_7)-(H_8)$,

$$\begin{aligned}
&\int_{\mathbb{R}^d}dz_1\int_{\mathbb{R}}K_1(z_1)K_2(v_1)\varphi\left(x+b_1z_1,\epsilon+m(x)+hv_1\right)dv_1 \\
&\quad-\varphi(x,\epsilon+m(x))-\frac{b_1^2}{2}\frac{\partial^2\varphi(x,\epsilon+m(x))}{\partial^2x}\int zK_0(z)z^\top dz-\frac{h^2}{2}\frac{\partial^2\varphi(x,\epsilon+m(x))}{\partial^2y}\int v^2K_1(v)dv \\
&= \frac{b_1^2}{2}\int\int z\left(\frac{\partial^2\varphi(x+\theta h_0z,\epsilon+m(x)+\theta h_1v)}{\partial^2x}-\frac{\partial^2\varphi(x,\epsilon+m(x))}{\partial^2x}\right)z^\top K_1(z)K_2(v)dzdv \\
&\quad+b_1h\int\int v\left(\frac{\partial^2\varphi(x+\theta b_1z,\epsilon+m(x)+\theta b_1v)}{\partial x\partial y}-\frac{\partial^2\varphi(x,\epsilon+m(x))}{\partial x\partial y}\right)z^\top K_1(z)K_2(v)dzdv \\
&\quad+\frac{h^2}{2}\int\int\left(\frac{\partial^2\varphi(x+\theta b_1z,\epsilon+m(x)+\theta b_1v)}{\partial^2y}-\frac{\partial^2\varphi(x,\epsilon+m(x))}{\partial^2y}\right)v^2K_1(z)K_2(v)dzdv.
\end{aligned}$$

Hence by the Lebesgue Dominated Convergence Theorem, we have, using (4.6.6) and (4.2.4),

$$\begin{aligned}
&\mathbb{E}\widetilde{f}_{2n}(\epsilon)-\frac{b_1^2}{2}\int \mathbb{1}\left(x\in\mathcal{X}\right)\frac{\partial^2\varphi(x,\epsilon+m(x))}{\partial^2x}dx\int zK_1(z)z^\top dz \\
&\quad-\frac{h^2}{2}\int \mathbb{1}\left(x\in\mathcal{X}\right)\frac{\partial^2\varphi(x,\epsilon+m(x))}{\partial^2y}dx\int v^2K_2(v)dv \\
&= \int \mathbb{1}\left(x\in\mathcal{X}\right)\varphi\left(x,\epsilon+m(x)\right)dx+o\left(b_1^2+h^2\right) \\
&= f(\epsilon)+o\left(b_1^2+h^2\right), \quad (4.6.7)
\end{aligned}$$

so that

$$\mathbb{E}\widetilde{f}_{2n}(\epsilon)-f(\epsilon)=O\left(b_1^2+h^2\right).$$

Finally, combining this result with (4.6.5) and (4.6.4), we arrive at

$$\widetilde{f}_{2n}(\epsilon)-f(\epsilon)=O_{\mathbb{P}}\left(b_1^4+h^4+\frac{1}{nh}\right)^{1/2}.\square$$

Proof of Theorem 4.2

Observe that

$$\begin{aligned} RT_n(b_0, b_0, h) &= b_0^4 + \frac{1}{nh^3}\left(b_0^4 + \frac{1}{nb_0^d}\right) + \frac{1}{n} + \frac{1}{nh^5}\left(b_0^4 + \frac{1}{nb_0^d}\right)^2 \\ &\quad + \frac{1}{n^2 b_0^d h^3} + \frac{1}{h^2}\left(b_0^4 + \frac{1}{nb_0^d}\right)^3 + \frac{b_0^d}{h^7}\left(b_0^4 + \frac{1}{nb_0^d}\right)^3. \end{aligned}$$

and note that

$$\left(\frac{1}{n^2h^3}\right)^{\frac{1}{d+4}} = \max\left\{\left(\frac{1}{n^2h^3}\right)^{\frac{1}{d+4}}, \left(\frac{1}{n^3h^7}\right)^{\frac{1}{2d+4}}\right\}$$

if and only if $n^{4-d}h^{d+16} \to \infty$. To find the order of b_0^*, we shall deal with the cases $nb_0^{d+4} \to \infty$ and $nb_0^{d+4} = O(1)$.

First assume that $nb_0^{d+4} \to \infty$. More precisely, we suppose that b_0 is in $\left[(u_n/n)^{1/(d+4)}, \infty\right)$, where $u_n \to \infty$. Since $1/(nb_0^d) = O(b_0^4)$ for all these b_0, we have

$$\left(b_0^4 + \frac{1}{nb_0^d}\right) \asymp \left(b_0^4\right), \quad \frac{1}{n^2 b_0^d h^3} = O\left(\frac{b_0^4}{nh^3}\right).$$

Hence the order of b_0^* is computed by minimizing the function

$$b_0 \to b_0^4 + \frac{b_0^4}{nh^3} + \frac{1}{n} + \frac{1}{nh^5}\left(b_0^4\right)^2 + \frac{1}{h^2}\left(b_0^4\right)^3 + \frac{b_0^d}{h^7}\left(b_0^4\right)^3.$$

Since this function is increasing with b_0, the minimum of $RT_n(\cdot, \cdot, h)$ is achieved for $b_{0*} = (u_n/n)^{1/(d+4)}$. We shall show later on that this choice of b_{0*} is irrelevant compared to the one arising when $nb_0^{d+4} = O(1)$.

Consider now the case $nb_0^{d+4} = O(1)$ i.e $b_0^4 = O\left(1/(nb_0^d)\right)$. This gives, since nb_0^{2d} diverges under (H_9), using $b_0^d/(nb_0^{2d})^p = O(b_0^{2p})$, $p = 2$,

$$\frac{1}{nh^3}\left(b_0^4 + \frac{1}{nb_0^d}\right) \asymp \frac{1}{n^2 b_0^d h^3}, \quad \frac{1}{nh^5}\left(b_0^4 + \frac{1}{nb_0^d}\right)^2 = O\left(\frac{1}{n^3 b_0^{2d} h^7}\right),$$

$$\frac{1}{h^2}\left(b_0^4 + \frac{1}{nb_0^d}\right)^3 = O\left(\frac{1}{n^2 b_0^d h^3}\right) \text{ and } \frac{b_0^d}{h^7}\left(b_0^4 + \frac{1}{nb_0^d}\right)^3 \asymp \left(\frac{1}{n^3 b_0^{2d} h^7}\right).$$

Moreover if $nb_0^d h^4 \to \infty$, we have

$$\frac{1}{n^3 b_0^{2d} h^7} = O\left(\frac{1}{n^2 b_0^d h^3}\right), \quad RT_n(b_0, b_0, h) = b_0^4 + \frac{1}{n^2 b_0^d h^3} + \frac{1}{n}.$$

Hence in this case, the order of b_0^* is obtained by finding the minimum of the function $b_0^4 + \left(1/n^2 b_0^d h^3\right) + (1/n)$. The minimization of this function gives a solution b_0 such that

$$b_0 \asymp \left(\frac{1}{n^2h^3}\right)^{\frac{1}{d+4}}, \quad RT_n(b_0, b_0, h) \asymp \frac{1}{n} + \left(\frac{1}{n^2h^3}\right)^{\frac{4}{d+4}}.$$

This value satisfies the constraints $nb_0^{d+4} = O(1)$ and $nb_0^d h^4 \to \infty$ when $n^{4-d}h^{d+16} \to \infty$.

If now $nb_0^{d+4} = O(1)$ but $nb_0^d h^4 = O(1)$, we have

$$\frac{1}{n^2 b_0^d h^3} = O\left(\frac{1}{n^3 b_0^{2d} h^7}\right), \quad RT_n(b_0, b_0, h) = b_0^4 + \frac{1}{n^3 b_0^{2d} h^7} + \frac{1}{n}.$$

In this case, the order of b_0^* is achieved by minimizing the function $b_0^4 + \left(1/n^3 b_0^{2d} h^7\right) + (1/n)$, for which the solution b_0 verifies

$$b_0 \asymp \left(\frac{1}{n^3 h^7}\right)^{\frac{1}{2d+4}}, \quad RT_n(b_0, b_0, h) \asymp \frac{1}{n} + \left(\frac{1}{n^3 h^7}\right)^{\frac{4}{2d+4}}.$$

This solution fulfills the constraint $nb_0^d h^4 = O(1)$ when $n^{4-d} h^{d+16} = O(1)$. Hence we can conclude that for $b_0^4 = O\left(1/(nb_0^d)\right)$, the bandwidth b_0^* satisfies

$$b_0^* \asymp \max\left\{\left(\frac{1}{n^2 h^3}\right)^{\frac{1}{d+4}}, \left(\frac{1}{n^3 h^7}\right)^{\frac{1}{2d+4}}\right\},$$

which leads to

$$RT_n\left(b_0^*, b_0^*, h\right) \asymp \frac{1}{n} + \max\left\{\left(\frac{1}{n^2 h^3}\right)^{\frac{4}{d+4}}, \left(\frac{1}{n^3 h^7}\right)^{\frac{4}{2d+4}}\right\}.$$

We need now to compare the solution b_0^* to the candidate $b_{0*} = (u_n/n)^{1/(d+4)}$ obtained when $nb_0^{d+4} \to \infty$. For this, we must do a comparison between the orders of $RT_n(b_0^*, b_0^*, h)$ and $RT_n(b_{0*}, b_{0*}, h)$. Since $RT_n(b_0, b_0, h) \geq b_0^4$, we have $RT_n(b_{0*}, b_{0*}, h) \geq (u_n/n)^{4/(d+4)}$, so that, for n large enough,

$$\begin{aligned}
\frac{RT_n(b_0^*, b_0^*, h)}{RT_n(b_{0*}, b_{0*}, h)} &\leq C\left[\left(\frac{1}{n^2 h^3}\right)^{\frac{1}{d+4}} + \left(\frac{1}{n^3 h^7}\right)^{\frac{4}{2d+4}}\right]\left(\frac{n}{u_n}\right)^{\frac{4}{d+4}} \\
&= o(1) + O\left(\frac{1}{u_n}\right)^{\frac{4}{d+4}}\left(\frac{1}{nh^{\frac{7(d+4)}{d+8}}}\right)^{\frac{4(d+8)}{(2d+4)(d+4)}} = o(1),
\end{aligned}$$

using $u_n \to \infty$ and $n^{(d+8)} h^{7(d+4)} \to \infty$ by (H_{10}). This shows that $RT_n(b_0^*, b_0^*, h) \leq RT_n(b_{0*}, b_{0^*}, h)$ for n large enough. This ends the proof of the Theorem, since b_0^* is the best candidate for the minimization of $RT_n(\cdot, \cdot, h)$. □

Proof of Theorem 4.3

The proof is the same as the one of Theorem 3.3 in Chapter 3. □

Proof of Theorem 4.4

The proof of the Theorem is based on the following Lemma.

Lemma 4.6. *Define*

$$\widetilde{\widetilde{f}}_{in}(\epsilon) = \frac{1}{b_1^d h} \int \mathbb{1}\left(x \in \mathcal{X}\right) K_1\left(\frac{X_i - x}{b_1}\right) K_2\left(\frac{Y_i - \epsilon - m(x)}{b_1}\right) dx.$$

Then, under (H_1), $(H_6)-(H_8)$, *we have, for* b_1 *and* h *and going to* 0 *and for some constant* $C>0$,

$$
\begin{aligned}
\mathbb{E}\widetilde{\widetilde{f}}_{in}(\epsilon) &= f(\epsilon)+\frac{b_1^2}{2}\int \mathbb{1}\left(x\in\mathcal{X}\right)\frac{\partial^2\varphi(x,\epsilon+m(x))}{\partial^2 x}dx\int zK_1(z)z^\top dz \\
&\quad +\frac{b_1^2}{2}\int \mathbb{1}\left(x\in\mathcal{X}\right)\frac{\partial^2\varphi(x,\epsilon+m(x))}{\partial^2 y}dx\int v^2K_2(v)dv+o\left(b_1^2+h^2\right),\\
\operatorname{Var}\left(\widetilde{\widetilde{f}}_{in}(\epsilon)\right) &= \frac{f(\epsilon)}{h}\int K_2^2(v)dv+o\left(\frac{1}{h}\right),\\
\mathbb{E}\left|\widetilde{\widetilde{f}}_{in}(\epsilon)-\mathbb{E}\widetilde{\widetilde{f}}_{in}(\epsilon)\right|^3 &\le \frac{Cf(\epsilon)}{h^2}\int\int\left|K_1(z_1)K_2(v_1)\right|^3 z_1dv_1+o\left(\frac{1}{h^2}\right).
\end{aligned}
$$

This Lemma is proved in Appendix B.

Let now turn to the proof of the Theorem 4.4. Observe that

$$
\widehat{f}_{2n}(\epsilon)-\mathbb{E}\widetilde{f}_{2n}(\epsilon)=\left(\widetilde{f}_{2n}(\epsilon)-\mathbb{E}\widetilde{f}_{2n}(\epsilon)\right)+\left(\widehat{f}_{2n}(\epsilon)-\widetilde{f}_{2n}(\epsilon)\right). \tag{4.6.8}
$$

Let now $\widetilde{\widetilde{f}}_{in}(\epsilon)$ be as in Lemma 4.6, and note that $\widetilde{f}_{2n}(\epsilon)=(1/n)\sum_{i=1}^n\widetilde{\widetilde{f}}_{in}(\epsilon)$. The second and the third claims in Lemma 4.6 yield, since h goes to 0 under (H_{10}),

$$
\frac{\sum_{i=1}^n\mathbb{E}\left|\widetilde{\widetilde{f}}_{in}(\epsilon)-\mathbb{E}\widetilde{\widetilde{f}}_{in}(\epsilon)\right|^3}{\left(\sum_{i=1}^n\operatorname{Var}\widetilde{\widetilde{f}}_{in}(\epsilon)\right)^3}\le\frac{\frac{Cnf(\epsilon)}{h^2}\int\int\left|K_1(z_1)K_2(v_1)\right|^3z_1dv_1+o\left(\frac{1}{h^2}\right)}{\left(\frac{nf(\epsilon)}{h}\int K_2^2(v)dv+o\left(\frac{n}{h}\right)\right)^3}=O(h)=o(1).
$$

Hence the Lyapounov Central Limit Theorem (Billingsley 1968, Theorem 7.3) gives, since nh diverges under (H_{10}),

$$
\frac{\widetilde{f}_{2n}(\epsilon)-\mathbb{E}\widetilde{f}_{2n}(\epsilon)}{\sqrt{\operatorname{Var}\widetilde{f}_{2n}(\epsilon)}}=\frac{\widetilde{f}_{2n}(\epsilon)-\mathbb{E}\widetilde{f}_{2n}(\epsilon)}{\sqrt{\frac{\operatorname{Var}\widetilde{\widetilde{f}}_{in}(\epsilon)}{n}}}\xrightarrow{d}\mathcal{N}\left(0,1\right),
$$

which yields, using the second result in Lemma 4.6,

$$
\sqrt{nh}\left(\widetilde{f}_{2n}(\epsilon)-\mathbb{E}\widetilde{f}_{2n}(\epsilon)\right)\xrightarrow{d}\mathcal{N}\left(0,f(\epsilon)\int K_2^2(v)dv\right). \tag{4.6.9}
$$

Observe now that Lemma 4.1 gives, for $b_1=b_0$,

$$
\begin{aligned}
\widehat{f}_{2n}(\epsilon)-\widetilde{f}_{2n}(\epsilon) &= O_{\mathbb{P}}\left[b_0^4+\frac{1}{nh^3}\left(b_0^4+\frac{1}{nb_0^d}\right)+\frac{1}{n}+\frac{1}{nh^5}\left(b_0^4+\frac{1}{nb_0^d}\right)^2\right]^{1/2}\\
&\quad +O_{\mathbb{P}}\left[\frac{1}{n^2b_0^dh^3}+\frac{1}{h^2}\left(b_0^4+\frac{1}{nb_0^d}\right)^3+\frac{b_0^d}{h^7}\left(b_0^4+\frac{1}{nb_0^d}\right)^3\right]^{1/2}.
\end{aligned}
$$

Moreover, since by Assumption (H_{11}) we have $nb_0^{d+4}=O(1)$, this ensures that $nb_0^{2d}\to\infty$ under (H_9), using $b_0^d/(nb_0^{2d})^p=O(b_0^{2p})$, $p=2$. Therefore

$$
\frac{1}{nh^3}\left(b_0^4+\frac{1}{nb_0^d}\right)\asymp\frac{1}{n^2b_0^dh^3},\quad\frac{1}{nh^5}\left(b_0^4+\frac{1}{nb_0^d}\right)^2=O\left(\frac{1}{n^3b_0^{2d}h^7}\right),
$$

$$\frac{1}{h^2}\left(b_0^4+\frac{1}{nb_0^d}\right)^3=O\left(\frac{1}{n^2b_0^dh^3}\right) \text{ and } \frac{b_0^d}{h^7}\left(b_0^4+\frac{1}{nb_0^d}\right)^3\asymp\left(\frac{1}{n^3b_0^{2d}h^7}\right).$$

Hence, for b_0 and h going to, it follows that

$$\sqrt{nh}\left(\widehat{f}_{2n}(\epsilon)-\widetilde{f}_{2n}(\epsilon)\right)\asymp O_{\mathbb{P}}\left[nh\left(b_0^4+\frac{1}{n^2b_0^dh^3}+\frac{1}{n}+\frac{1}{n^3b_0^{2d}b_1^7}\right)\right]^{1/2}=o_{\mathbb{P}}(1),$$

since $nb_0^4h=o(1)$ and $nb_0^dh^3\to\infty$ by Assumption (H_{11}). Hence from (4.6.9) and (4.6.8), we deduce

$$\sqrt{nh}\left(\widehat{f}_{2n}(\epsilon)-\mathbb{E}\widetilde{f}_{2n}(\epsilon)\right)\xrightarrow{d}\mathcal{N}\left(0,f(\epsilon)\int K_2^2(v)dv\right).$$

This proves the Theorem, since the first result of Lemma 4.6 gives for $b_1=b_0$,

$$\begin{aligned}\mathbb{E}\widetilde{f}_{2n}(\epsilon) \quad = \quad & f(\epsilon)+\frac{b_0^2}{2}\int\mathbb{1}\left(x\in\mathcal{X}\right)\frac{\partial^2\varphi(x,\epsilon+m(x))}{\partial^2x}dx\int zK_1(z)z^\top dz\\ &+\frac{h^2}{2}\int\mathbb{1}\left(x\in\mathcal{X}\right)\frac{\partial^2\varphi(x,\epsilon+m(x))}{\partial^2y}dx\int v^2K_2(v)dv+o\left(b_0^2+h^2\right):=\overline{f}_{2n}(\epsilon).\square\end{aligned}$$

Appendix B : Proof of Lemmas 4.3-4.6

Intermediate results for Lemmas 4.3-4.5

Lemma 4.7. *If $(H_1)-(H_2)$, (H_7) and (H_9) are satisfied, we have*

$$\begin{aligned}\sup_{x\in\mathcal{X}}|\widehat{g}_n(x)-g(x)| &= O_{\mathbb{P}}\left(b_0^4+\frac{\ln n}{nb_0^d}\right)^{1/2},\\ \sup_{x\in\mathcal{X}}\left|\frac{1}{\widehat{g}_n(x)}-\frac{1}{g(x)}\right| &= O_{\mathbb{P}}\left(b_0^4+\frac{\ln n}{nb_0^d}\right)^{1/2}.\end{aligned}$$

Lemma 4.8. *Let $\mathbb{E}_{in}[\cdot]$ be the conditional mean given $(X_1,\ldots,X_n,\varepsilon_k,k\neq i)$. Then if $(H_1)-(H_5)$, (H_8) and (H_{10}) hold, we have, for any integer $i\in[1,n]$, $p\in[0,6]$ and $y\in\mathbb{R}$,*

$$\left|\mathbb{E}_{in}\left[\varepsilon_i^pK_2^{(1)}\left(\frac{Y_i-y}{h}\right)\right]\right|\le Ch^2,\quad \left|\mathbb{E}_{in}\left[\varepsilon^pK_2^{(1)}\left(\frac{Y_i-y}{h}\right)^2\right]\right|\le Ch, \tag{B.1}$$

$$\left|\mathbb{E}_{in}\left[\varepsilon_i^pK_2^{(2)}\left(\frac{Y_i-y}{h}\right)\right]\right|\le Ch^3,\quad \left|\mathbb{E}_{in}\left[\varepsilon_i^pK_2^{(2)}\left(\frac{Y_i-y}{h}\right)^2\right]\right|\le Ch, \tag{B.2}$$

$$\left|\mathbb{E}_{in}\left[\varepsilon_i^pK_2^{(3)}\left(\frac{Y_i-y}{h}\right)\right]\right|\le Ch^3,\quad \left|\mathbb{E}_{in}\left[\varepsilon_i^pK_2^{(3)}\left(\frac{Y_i-y}{h}\right)^2\right]\right|\le Ch, \tag{B.3}$$

for some constant $C>0$.

Let $\mathbb{E}_n[\cdot]$ and $\mathrm{Var}_n[\cdot]$ be respectively the conditional mean and the conditional variance given $(X_1,\ldots,X_n)$, and denote $b_0\vee b_1=\max(b_0,b_1)$. In the following, S_n and T_n are defined as in Lemma 4.3. Then the following results are used in the proof of Lemmas 4.3, 4.4 and 4.5.

Lemma 4.9. *If $(H_1)-(H_{10})$ hold, then*

$$\mathbb{E}_n[S_n]=O_{\mathbb{P}}\left(b_0^2\right),\quad \mathbb{E}_n[T_n]=O_{\mathbb{P}}\left(b_0^4+\frac{1}{nb_0^d}\right).$$

Lemma 4.10. *Under $(H_1)-(H_{10})$, we have*

$$\begin{aligned}\mathrm{Var}_n[S_n] &= O_{\mathbb{P}}\left(b_0^d\vee b_1^d\right)\left[\frac{1}{nb_1^dh^3}\left(b_0^4+\frac{1}{nb_0^d}\right)+\frac{1}{nb_0^d}\right],\\ \mathrm{Var}_n[T_n] &= O_{\mathbb{P}}\left(b_0^d\vee b_1^d\right)\left[\frac{1}{nb_1^dh^5}\left(b_0^4+\frac{1}{nb_0^d}\right)^2+\frac{b_0^4}{nb_0^d}+\frac{1}{n^2b_0^{2d}h^3}\right].\end{aligned}$$

Lemma 4.11. *Define for all integer number p in $[1,3]$,*

$$U_n(x)=U_n(x;p)=\frac{1}{nb_1^dh^{p+1}}\sum_{i=1}^n\left(\widehat{m}_{in}(x)-m(x)\right)^pK_1\left(\frac{X_i-x}{b_1}\right)K_2^{(p)}\left(\frac{Y_i-\epsilon-m(x)}{h}\right),$$

and assume that (H_4) and (H_7) hold. Consider C large enough and any x_1, x_2 in $\mathcal{X}$ with $\|x_2-x_1\|\ge Cb_0\vee b_1$. Then $U_n(x_1)$ and $U_n(x_2)$ are independent given $X_1,\ldots,X_n$.

Lemma 4.12. *Set*

$$\begin{aligned}\beta_{in}(x) &= \frac{\sum_{1\leq j\neq i\leq n}\left(m(X_j)-m(x)\right)K_0\left(\frac{X_j-x}{b_0}\right)}{nb_0^d\widehat{g}_n(x)},\\ \Sigma_{in}(x) &= \frac{\sum_{1\leq j\neq i\leq n}\varepsilon_j K_0\left(\frac{X_j-x}{b_0}\right)}{nb_0^d\widehat{g}_n(x)}.\end{aligned}$$

Then under $(H_1)-(H_5)$ *and* $(H_7)-(A_9)$*, we have, for all integers* p_1 *and* p_2 *in* $[0,6]$,

$$\sum_{i=1}^n \int \mathbb{1}(x\in\mathcal{X})\left|\beta_{in}^{p_1}(x)K_1^{p_2}\left(\frac{X_i-x}{b_1}\right)\right|dx = O_{\mathbb{P}}\left(nb_1^d\right)\left(b_0^{2p_1}\right), \tag{B.4}$$

$$\sum_{i=1}^n \int \mathbb{1}(x\in\mathcal{X})\,\mathbb{E}_n\left|\Sigma_{in}^{p_1}(x)K_1^{p_2}\left(\frac{X_i-x}{b_1}\right)\right|dx = O_{\mathbb{P}}\left(\frac{nb_1^d}{(nb_0^d)^{p_1/2}}\right). \tag{B.5}$$

The proof of these lemmas are given in Appendix C.

Proof of Lemma 4.3

The proof follows directly from Lemmas 4.9 and 4.10. Indeed, since the Tchebychev inequality, which ensures that

$$A_n = O_{\mathbb{P}}\left(\mathbb{E}_n\left[A_n\right] + \mathrm{Var}_n^{1/2}\left(A_n\right)\right),$$

Lemmas 4.9 and 4.10 then give

$$\begin{aligned} S_n &= O_{\mathbb{P}}\left[b_0^4 + \left(b_0^d \vee b_1^d\right)\left(\frac{1}{nb_1^d h^3}\left(b_0^4 + \frac{1}{nb_0^d}\right) + \frac{1}{nb_0^d}\right)\right]^{1/2}, \\ T_n &= O_{\mathbb{P}}\left[\left(b_0^4 + \frac{1}{nb_0^d}\right)^2 + \left(b_0^d \vee b_1^d\right)\left(\frac{1}{nb_1^d h^5}\left(b_0^4 + \frac{1}{nb_0^d}\right)^2 + \frac{b_0^4}{nb_0^d} + \frac{1}{n^2 b_0^{2d} h^3}\right)\right]^{1/2}, \end{aligned}$$

which proves Lemma 4.3. □

Proof of Lemma 4.4

Set

$$R_n = \int \mathbb{1}\left(x \in \mathcal{X}\right) R_n(x) dx.$$

The proof of the Lemma proceeds by computing the conditional mean and the conditional variance of R_n. For the conditional mean, define

$$\begin{aligned} I_{in}(x) &= \int_0^1 (1-u)^2 K_2^{(3)}\left(\frac{Y_i - \epsilon - m(x) - u\left(\widehat{m}_{in}(x) - m(x)\right)}{h}\right) du, \\ R_{in}(x) &= \frac{1}{nb_1^d h^4} K_1\left(\frac{X_i - x}{b_1}\right)\left(\widehat{m}_{in}(x) - m(x)\right)^3 I_{in}(x). \end{aligned}$$

This gives

$$\mathbb{E}_n\left[R_n\right] = \sum_{i=1}^n \int \mathbb{1}\left(x \in \mathcal{X}\right) \mathbb{E}_n\left[R_{in}(x)\right] dx, \tag{B.6}$$

where

$$\mathbb{E}_n\left[R_{in}(x)\right] = \frac{1}{nb_1^d h^4} K_1\left(\frac{X_i - x}{b_1}\right) \mathbb{E}_n\left[\left(\widehat{m}_{in}(x) - m(x)\right)^3 \mathbb{E}_{in}\left[I_{in}(x)\right]\right].$$

Moreover, since by Lemma 4.8-(B.3) we have

$$\left|\mathbb{E}_{in}\left[I_{in}(x)\right]\right| = \left|\int_0^1 (1-u)^2 \mathbb{E}_{in}\left[K_2^{(3)}\left(\frac{Y_i - \epsilon - m(x) - u\left(\widehat{m}_{in}(x) - m(x)\right)}{h}\right)\right] du\right| \leq Ch^3,$$

it then follows that, setting $p_1 = 3$ and $p_2 = 1$ Lemma 4.12,

$$\begin{aligned}
&|\mathbb{E}_n[R_n]| \\
\leq\ & \frac{Ch^3}{nb_1^d h^4} \sum_{i=1}^n \int \mathbb{1}(x \in \mathcal{X}) \mathbb{E}_n \left| (\widehat{m}_{in}(x) - m(x))^3 K_1 \left(\frac{X_i - x}{b_1} \right) \right| dx \\
\leq\ & \frac{C}{nb_1^d h} \sum_{i=1}^n \int \mathbb{1}(x \in \mathcal{X}) \left| \beta_{in}^3(x) K_1 \left(\frac{X_i - x}{b_1} \right) \right| dx \\
& + \frac{C}{nb_1^d h} \sum_{i=1}^n \int \mathbb{1}(x \in \mathcal{X}) \mathbb{E}_n \left| \Sigma_{in}^3(x) K_1 \left(\frac{X_i - x}{b_1} \right) \right| dx \\
=\ & O_{\mathbb{P}} \left[\frac{1}{h^2} \left(b_0^4 + \frac{1}{nb_0^d} \right)^3 \right]^{1/2}. \qquad \text{(B.7)}
\end{aligned}$$

Consider now the conditional variance of R_n. Let C large enough and consider x_1, x_2 in $\mathcal{X}$ with $\|x_2 - x_1\| \geq Cb_0 \vee b_1$. Then given $X_1, \ldots, X_n$ and under (H_4), there exists two functions Φ_{1n} and Φ_{2n} such that

$$R_n(x_1) = \Phi_{1n}(\varepsilon_i, i \in I_1) \text{ and } R_n(x_2) = \Phi_{2n}(\varepsilon_i, i \in I_2),$$

with an empty $I_1 \cap I_2$, since the Kernel functions K_0 and K_1 are compactly supported. Hence $R_n(x_1)$ and $R_n(x_2)$ are independent given $X_1, \ldots, X_n$, provided that $\|x_2 - x_1\| \geq Cb_0 \vee b_1$, for C sufficiently large. Therefore

$$\begin{aligned}
&\operatorname{Var}_n(R_n) \\
=\ & \operatorname{Var}_n \left(\int \mathbb{1}(x \in \mathcal{X}) R_n(x)\, dx \right) = \int\int \mathbb{1}\left((x_1, x_2) \in \mathcal{X}^2\right) \operatorname{Cov}_n(R_n(x_1), R_n(x_2))\, dx_1 dx_2 \\
\leq\ & \int\int \mathbb{1}\left((x_1, x_2) \in \mathcal{X}^2, \|x_2 - x_1\| \leq Cb_0 \vee b_1\right) \operatorname{Var}_n^{1/2}(R_n(x_1)) \operatorname{Var}_n^{1/2}(R_n(x_2))\, dx_1 dx_2 \\
\leq\ & \frac{1}{2} \int\int \mathbb{1}\left((x_1, x_2) \in \mathcal{X}^2, \|x_2 - x_1\| \leq Cb_0 \vee b_1\right) \{\operatorname{Var}_n(R_n(x_1)) + \operatorname{Var}_n(R_n(x_2))\}\, dx_1 dx_2 \\
\leq\ & C(b_0 \vee b_1)^d \int \mathbb{1}(x \in \mathcal{X}) \operatorname{Var}_n(R_n(x))\, dx, \qquad \text{(B.8)}
\end{aligned}$$

where

$$\begin{aligned}
&\int \mathbb{1}(x \in \mathcal{X}) \operatorname{Var}_n(R_n(x))\, dx \\
&= \sum_{i=1}^n \int \mathbb{1}(x \in \mathcal{X}) \operatorname{Var}_n(R_{in}(x))\, dx \\
&\quad + \sum_{1 \leq i_1 \neq i_2 \leq n} \int \mathbb{1}(x \in \mathcal{X}) \operatorname{Cov}_n(R_{i_1 n}(x), R_{i_2 n}(x))\, dx. \qquad \text{(B.9)}
\end{aligned}$$

For the conditional variances in (B.9), we have

$$\operatorname{Var}_n(R_{in}(x)) \leq \frac{1}{(nb_1^d h^4)^2} K_0^2 \left(\frac{X_i - x}{b_1} \right) \mathbb{E}_n \left[(\widehat{m}_{in}(x) - m(x))^6 I_{in}^2(x) \right]$$

with, applying Lemma 4.8-(B.3),

$$\begin{aligned}
\mathbb{E}_n\left[(\widehat{m}_{in}(x)-m(x))^6\, I_{in}^2(x)\right] &= \mathbb{E}_n\left[(\widehat{m}_{in}(x)-m(x))^6\, \mathbb{E}_{in}\left[I_{in}^2(x)\right]\right]\\
&\leq C\mathbb{E}_n\left[(\widehat{m}_{in}(x)-m(x))^6 \sup_{y\in\mathbb{R}} \mathbb{E}_{in}\left[K_2^{(3)}\left(\frac{Y_i-y}{h}\right)^2\right]\right]du\\
&\leq Ch\mathbb{E}_n\left[(\widehat{m}_{in}(x)-m(x))^6\right].
\end{aligned}$$

Hence from this result and Lemma 4.12, we deduce

$$\begin{aligned}
&\sum_{i=1}^n \int \mathbb{1}\left(x\in\mathcal{X}\right)\mathrm{Var}_n\left(R_{in}\left(x\right)\right)dx\\
&\leq \frac{Ch}{(nb_1^dh^4)^2}\sum_{i=1}^n\int \mathbb{1}\left(x\in\mathcal{X}\right)\mathbb{E}_n\left[(\widehat{m}_{in}(x)-m(x))^6\right]K_1^2\left(\frac{X_i-x}{b_1}\right)dx\\
&\leq \frac{Ch}{(nb_1^dh^4)^2}\sum_{i=1}^n\int \mathbb{1}\left(x\in\mathcal{X}\right)\left(\beta_{in}^6(x)+\mathbb{E}\left[\Sigma_{in}^6(x)\right]\right)K_1^2\left(\frac{X_i-x}{b_1}\right)dx\\
&= \frac{O_{\mathbb{P}}\left(nb_1^dh\right)}{(nb_1^dh^4)^2}\left(b_0^4+\frac{1}{nb_0^d}\right)^3. \qquad \text{(B.10)}
\end{aligned}$$

Let now turn to the sum of the conditional covariances in (B.9). We have

$$\left|\mathrm{Cov}_n\left(R_{i_1n}(x),R_{i_2n}(x)\right)\right| \leq \mathrm{Var}_n^{1/2}\left(R_{i_1n}(x)\right)\mathrm{Var}_n^{1/2}\left(R_{i_2n}(x)\right),$$

where

$$\mathrm{Var}_n\left(R_{i_1n}(x)\right) \leq \frac{Ch}{(nb_1^dh^4)^2}\mathbb{E}_n\left[(\widehat{m}_{i_1n}(x)-m(x))^6\right]K_1^2\left(\frac{X_{i_1}-x}{b_1}\right).$$

Hence

$$\begin{aligned}
&O_{\mathbb{P}}\left(\frac{(nb_1^dh^4)^2}{h}\right)\sum_{1\leq i_1\neq i_2\leq n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\left|\mathrm{Cov}_n\left(R_{i_1n}(x),R_{i_2n}(x)\right)\right|dx\\
&= \sum_{1\leq i_1\neq i_2\leq n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\mathbb{E}_n^{1/2}\left[(\widehat{m}_{i_1n}(x)-m(x))^6\right]\mathbb{E}_n^{1/2}\left[(\widehat{m}_{i_2n}(x)-m(x))^6\right]\\
&\qquad\times\left|K_1\left(\frac{X_{i_1}-x}{b_1}\right)K_1\left(\frac{X_{i_2}-x}{b_1}\right)\right|dx\\
&\leq \sum_{1\leq i_1\neq i_2\leq n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\mathbb{E}_n\left[(\widehat{m}_{i_1n}(x)-m(x))^6\right]\left|K_1\left(\frac{X_{i_1}-x}{b_1}\right)K_1\left(\frac{X_{i_2}-x}{b_1}\right)\right|dx\\
&\quad+\sum_{1\leq i_1\neq i_2\leq n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\mathbb{E}_n\left[(\widehat{m}_{i_2n}(x)-m(x))^6\right]\left|K_1\left(\frac{X_{i_1}-x}{b_1}\right)K_1\left(\frac{X_{i_2}-x}{b_1}\right)\right|dx. \qquad \text{(B.11)}
\end{aligned}$$

Moreover, under (H_7), the change of variable $x = u + b_1 X_{i_2}$ and Lemma 4.12 give

$$\begin{aligned}
&\sum_{1\le i_1\neq i_2\le n} \int \mathbb{1}\left(x\in\mathcal{X}\right)\mathbb{E}_n\left[\left(\widehat{m}_{i_1n}(x)-m(x)\right)^6\right]\left|K_1\left(\frac{X_{i_1}-x}{b_1}\right)K_1\left(\frac{X_{i_2}-x}{b_1}\right)\right|dx\\
=\ & b_1^d \sum_{1\le i_1\neq i_2\le n} \int \mathbb{1}\left(u+b_1X_{i_2}\in\mathcal{X}\right)\mathbb{E}_n\left[\left(\widehat{m}_{i_1n}(u+b_1X_{i_2})-m(u+b_1X_{i_2})\right)^6\right]\\
&\qquad\times\left|K_1\left(u\right)K_1\left(\frac{X_{i_2}-u+b_1X_{i_2}}{b_1}\right)\right|du\\
=\ & O_{\mathbb{P}}\left(nb_1^d\right)\sum_{i=1}^n \int \mathbb{1}\left(x\in\mathcal{X}\right)\mathbb{E}_n\left[\left(\widehat{m}_{in}(x)-m(x)\right)^6\right]\left|K_1\left(\frac{X_i-x}{b_1}\right)\right|dx\\
=\ & O_{\mathbb{P}}\left(nb_1^d\right)\sum_{i=1}^n \int \mathbb{1}\left(x\in\mathcal{X}\right)\left(\beta_{in}^6(x)+\mathbb{E}\left[\Sigma_{in}^6(x)\right]\right)\left|K_1\left(\frac{X_i-x}{b_1}\right)\right|dx\\
=\ & O_{\mathbb{P}}\left(n^2b_1^{2d}\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^3.
\end{aligned}$$

Therefore collecting this result and (B.11), we arrive at

$$\sum_{1\le i_1\neq i_2\le n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\mathrm{Cov}_n\left(R_{i_1n}(x),R_{i_2n}(x)\right)dx = \frac{O_{\mathbb{P}}\left(n^2b_1^{2d}h\right)}{(nb_1^dh^4)^2}\left(b_0^4+\frac{1}{nb_0^d}\right)^3.$$

Substituting this order and (B.10) in (B.9), it follows, since $nb_1^d\to\infty$ under (H_{10}),

$$\begin{aligned}
\int \mathbb{1}\left(x\in\mathcal{X}\right)\mathrm{Var}_n\left(R_n\left(x\right)\right)dx &= O_{\mathbb{P}}\left[\frac{nb_1^dh}{(nb_1^dh^4)^2}+\frac{n^2b_1^{2d}h}{(nb_1^dh^4)^2}\right]\left(b_0^4+\frac{1}{nb_0^d}\right)^3\\
&= O_{\mathbb{P}}\left(\frac{1}{h^7}\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^3.
\end{aligned}$$

Hence by (B.8), (B.7) and the Tchebychev inequality, we have

$$R_n = O_{\mathbb{P}}\left[\frac{1}{h^2}\left(b_0^4+\frac{1}{nb_0^d}\right)^3+\frac{b_0^d\vee b_1^d}{h^7}\left(b_0^4+\frac{1}{nb_0^d}\right)^3\right]^{1/2},$$

which proves the validity of the Lemma. □

Proof of Lemma 4.5

Set

$$P_n = \int \mathbb{1}\left(x\in\mathcal{X}\right)P_n(x)dx.$$

The proof of the Lemma follows by computing the conditional mean and the conditional variance of P_n. For the conditional mean, define

$$\begin{aligned}
\widehat{I}_{in}(x) &= \int_0^1 K_2^{(1)}\left(\frac{Y_i-\epsilon+\widehat{m}_{in}(x)-t\left(\widehat{m}_n(x)-\widehat{m}_{in}(x)\right)}{b_1}\right)dt,\\
P_{in}(x) &= \frac{1}{nb_1^dh^2}\left(\widehat{m}_n(x)-\widehat{m}_{in}(x)\right)K_1\left(\frac{X_i-x}{b_1}\right)\widehat{I}_{in}(x).
\end{aligned}$$

Since

$$\widehat{m}_n(x)-\widehat{m}_{in}(x) = \frac{Y_i}{nb_0^d\widehat{g}_n(x)}K_0\left(\frac{X_i-x}{b_0}\right),$$

and that K_0 is bounded under (H_7), Lemma 4.7 gives

$$\begin{aligned}\mathbb{E}_n\left[P_n\right] &= \sum_{i=1}^{n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\mathbb{E}_n\left[P_{in}(x)\right]dx \\ &= O_{\mathbb{P}}\left(\frac{1}{nb_0^d}\right)\left[\frac{1}{nb_1^d h^2}\sum_{i=1}^{n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\left|K_1\left(\frac{X_i-x}{b_1}\right)\mathbb{E}_n\left[Y_i\widehat{I}_{in}(x)\right]\right|dx\right]. \qquad \text{(B.12)}\end{aligned}$$

Moreover, observe that for any $y\in\mathbb{R}$,

$$\mathbb{E}_{in}\left[Y_iK_2^{(1)}\left(\frac{Y_i-y}{h}\right)\right] = m(X_i)\mathbb{E}_{in}\left[K_2^{(1)}\left(\frac{Y_i-y}{h}\right)\right]+\mathbb{E}_{in}\left[\varepsilon_iK_2^{(1)}\left(\frac{Y_i-y}{h}\right)\right].$$

Therefore, since $m(\cdot)$ is continuous on the compact support $\mathcal{X}$ of the X_i's, Lemma 4.8-(B.1) yields

$$\sup_{y\in\mathbb{R}}\left|\mathbb{E}_{in}\left[Y_iK_2^{(1)}\left(\frac{Y_i-y}{h}\right)\right]\right| \le Ch^2,$$

uniformly for $i\in[1,n]$. Hence conditioning with respect to $(X_1,\ldots,X_n,\varepsilon_k)$ yields that

$$\left|\mathbb{E}_n\left[Y_i\widehat{I}_{in}(x)\right]\right| \le \left|\sup_{y\in\mathbb{R}}\int \mathbb{E}_{in}\left[Y_iK_2^{(1)}\left(\frac{Y_i-y}{b_1}\right)\right]dy\right| \le Ch^2,$$

for all i and x. Combining this result with (B.12), we arrive at

$$\begin{aligned}\mathbb{E}_n\left[P_n\right] &= O_{\mathbb{P}}\left(\frac{1}{nb_0^d}\right)\left[\frac{1}{nb_1^d}\sum_{i=1}^{n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\left|K_1\left(\frac{X_i-x}{b_1}\right)\right|dx\right] \\ &= O_{\mathbb{P}}\left(\frac{1}{nb_0^d}\right). \qquad \text{(B.13)}\end{aligned}$$

Let now consider the conditional variance of P_n. Since

$$P_{in}(x) = \frac{1}{nb_1^dh^2}\left[\frac{Y_i}{nb_0^d\widehat{g}_n(x)}K_0\left(\frac{X_i-x}{b_0}\right)\right]K_1\left(\frac{X_i-x}{b_1}\right)\widehat{I}_{in}(x),$$

and that $K_0(\cdot)$ and $K_1(\cdot)$ have compact supports under (H_7) and (H_8), it is shown that $P_n\left(x_1\right)$ and $P_n\left(x_2\right)$ are independent given $X_1,\ldots,X_n$, provided that $\|x_2-x_1\|\ge Cb_0\vee b_1$, for C large enough. Hence arguing as for (B.8) gives

$$\mathrm{Var}_n\left(P_n\right) \le C\left(b_0^d\vee b_1^d\right)\int \mathbb{1}\left(x\in\mathcal{X}\right)\mathrm{Var}_n\left(P_n\left(x\right)\right)dx, \qquad \text{(B.14)}$$

where

$$\begin{aligned}&\int \mathbb{1}\left(x\in\mathcal{X}\right)\mathrm{Var}_n\left(P_n\left(x\right)\right)dx \\ &\quad= \sum_{i=1}^{n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\mathrm{Var}_n\left(P_{in}\left(x\right)\right)dx \\ &\qquad+ \sum_{1\le i_1\ne i_2\le n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\mathrm{Cov}_n\left(P_{i_1n}(x),P_{i_2n}(x)\right)dx. \qquad \text{(B.15)}\end{aligned}$$

For the conditional variances in (B.15), first note that

$$\mathrm{Var}_n\left(P_{in}(x)\right) \le \frac{1}{(nb_1^dh^2)^2}K_1^2\left(\frac{X_i-x}{b_1}\right)\left[\frac{1}{(nb_0^d)^2\widehat{g}_n^2(x)}K_0^2\left(\frac{X_i-x}{b_0}\right)\mathbb{E}_n\left[\left(Y_i-m(x)\right)^2\widehat{I}_{in}^2(x)\right]\right]. \qquad \text{(B.16)}$$

Next, observe that for $X_i = z$ and $y \in \mathbb{R}$, and under (H_1), $(H_3) - (H_5)$ and (H_7), we have

$$\begin{aligned}\mathbb{E}_n\left[Y_i^2 K_2^{(1)}\left(\frac{Y_i - y}{h}\right)^2\right] &= \int (m(z)+e)^2 K_2^{(1)}\left(\frac{m(z)+e-y}{h}\right)^2 f(e)de \\ &\leq Ch,\end{aligned}$$

uniformly in x and i. From this result and the Hölder inequality, we deduce

$$\begin{aligned}&\mathbb{E}_n\left[Y_i^2 \widehat{I}_{in}^2(x)\right] \\ &\leq \int_0^1 \mathbb{E}_n\left[Y_i^2 K_2^{(1)}\left(\frac{Y_i - \epsilon + \widehat{m}_{in}(x) - t\left(\widehat{m}_n(x) - \widehat{m}_{in}(x)\right)}{h}\right)^2\right] dt \\ &\leq \sup_{y\in\mathbb{R}} \mathbb{E}_n\left[Y_i^2 K_2^{(1)}\left(\frac{Y_i - y}{h}\right)^2\right] \\ &\leq Ch.\end{aligned}$$

Hence by (B.16) and Lemma 4.7, we have, since $K_0(\cdot)$ is bounded under (H_7),

$$\operatorname{Var}_n\left(P_{in}(x)\right) \leq \frac{C}{(nb_1^d h^2)^2} \times \frac{h}{(nb_0^{2d})^2 \widehat{g}_n^2(x)} K_1^2\left(\frac{X_i - x}{b_1}\right),$$

so that

$$\begin{aligned}&\sum_{i=1}^n \int \mathbb{1}\left(x \in \mathcal{X}\right) \operatorname{Var}_n\left(P_{in}(x)\right) dx \\ &= O_{\mathbb{P}}\left(\frac{h}{(nb_1^d h^2)^2}\right)\left(\frac{1}{(nb_0^d)^2}\right) \sum_{i=1}^n \int \mathbb{1}\left(x \in \mathcal{X}\right) K_1^2\left(\frac{X_i - x}{b_1}\right) dx \\ &= O_{\mathbb{P}}\left(\frac{1}{nb_1^d h^3}\right)\left(\frac{1}{n^2 b_0^{2d}}\right). \qquad \text{(B.17)}\end{aligned}$$

Let now consider the sum of the conditional covariances in (B.15). We have, using the inequality above,

$$\begin{aligned}\left|\operatorname{Cov}_n\left(P_{i_1 n}(x), P_{i_2 n}(x)\right)\right| &\leq \operatorname{Var}_n^{1/2}\left(P_{i_1 n}(x)\right) \operatorname{Var}_n^{1/2}\left(P_{i_2 n}(x)\right) \\ &\leq \frac{C}{(nb_1^d h^2)^2} \times \frac{h}{(nb_0^{2d})^2 \widehat{g}_n^2(x)} \left|K_1\left(\frac{X_{i_1} - x}{b_1}\right) K_1\left(\frac{X_{i_2} - x}{b_1}\right)\right|.\end{aligned}$$

Hence from Lemma 4.7, we deduce

$$\begin{aligned}&\sum_{1 \leq i_1 \neq i_2 \leq n} \int \mathbb{1}\left(x \in \mathcal{X}\right) \left|\operatorname{Cov}_n\left(P_{i_1 n}(x), P_{i_2 n}(x)\right)\right| dx \\ &= O_{\mathbb{P}}\left(\frac{1}{(nb_1^d h^2)^2}\right)\left(\frac{h}{(nb_0^d)^2}\right) \sum_{1 \leq i_1 \neq i_2 \leq n} \int \mathbb{1}\left(x \in \mathcal{X}\right) \left|K_1\left(\frac{X_{i_1} - x}{b_1}\right) K_1\left(\frac{X_{i_2} - x}{b_1}\right)\right| dx \\ &= O_{\mathbb{P}}\left(\frac{1}{n^2 b_0^{2d} h^3}\right).\end{aligned}$$

Substituting this order and (B.17) in (B.15), and using (B.14), (B.13) and the Tchebychev inequality, we arrive at

$$\begin{aligned}P_n &= O_{\mathbb{P}}\left[\frac{1}{nb_0^d} + \left(b_0^d \vee b_1^d\right)^{1/2}\left(\frac{1}{nb_1^d h^3}\left(\frac{1}{n^2 b_0^{2d}}\right) + \frac{1}{n^2 b_0^{2d} h^3}\right)^{1/2}\right] \\ &= O_{\mathbb{P}}\left(\frac{1}{n^2 b_0^{2d}} + \frac{b_0^d \vee b_1^d}{n^2 b_0^{2d} h^3}\right)^{1/2}.\end{aligned}$$

This ends the proof of the Lemma. □

Proof of Lemma 4.6

The first equality of the lemma is given by (4.6.7), since $\widetilde{f}_{2n}(\epsilon) = (1/n)\sum_{i=1}^{n}\widetilde{\widetilde{f}}_{in}(\epsilon)$, so that

$$
\begin{aligned}
\mathbb{E}\widetilde{\widetilde{f}}_{in}(\epsilon) &= \mathbb{E}\widetilde{f}_{2n}(\epsilon) \\
&= f(\epsilon) + \frac{b_1^2}{2}\int \mathbb{1}\left(x \in \mathcal{X}\right) \frac{\partial^2 \varphi(x, \epsilon + m(x))}{\partial^2 x} dx \int z K_1(z) z^\top dz \\
&\quad + \frac{h^2}{2}\int \mathbb{1}\left(x \in \mathcal{X}\right) \frac{\partial^2 \varphi(x, \epsilon + m(x))}{\partial^2 y} dx \int v^2 K_2(v) dv + o\left(b_1^2 + h^2\right).
\end{aligned}
$$

For the second result of the Lemma, we have

$$
\begin{aligned}
\operatorname{Var}\left(\widetilde{\widetilde{f}}_{in}(\epsilon)\right) &= \mathbb{E}\left[\widehat{f}_{in}^2(\epsilon)\right] - \mathbb{E}^2\left[\widetilde{\widetilde{f}}_{in}(\epsilon)\right] \\
&= \frac{1}{b_1^{2d}h^2}\mathbb{E}\left[\left[\int \mathbb{1}\left(x \in \mathcal{X}\right) K_1\left(\frac{X_i - x}{b_1}\right) K_2\left(\frac{Y_i - \epsilon - m(x)}{b_1}\right) dx\right]^2\right] + O(1). \qquad \text{(B.18)}
\end{aligned}
$$

Observe now that the changes of variables $x = x_1 + b_1 z_1$ and $y_1 = \epsilon + m(x_1 + b_1 z_1) + b_1 v_1$ give

$$
\begin{aligned}
&\mathbb{E}\left[\left[\int \mathbb{1}\left(x \in \mathcal{X}\right) K_1\left(\frac{X_i - x}{b_1}\right) K_2\left(\frac{Y_i - \epsilon - m(x)}{b_1}\right) dx\right]^2\right] \\
&= \int dx_1 \int \left[\int \mathbb{1}\left(x \in \mathcal{X}\right) K_1\left(\frac{x_1 - x}{b_1}\right) K_2\left(\frac{y_1 - \epsilon - m(x)}{b_1}\right) dx\right]^2 \varphi\left(x_1, y_1\right) dy_1 \\
&= b_1^{2d} h \int dx_1 \int \left[K_2(v_1)\int \mathbb{1}\left(x_1 + b_1 z_1 \in \mathcal{X}\right) K_1(z_1) dz_1\right]^2 \varphi\left(x_1, \epsilon + m(x_1 + b_1 z_1) + b_1 v_1\right) dv_1.
\end{aligned}
\qquad \text{(B.19)}
$$

Moreover, note that under (H_3) and (H_6) we have

$$
\begin{aligned}
m\left(x_1 + b_1 z_1\right) &= m(x_1) + b_1 z_1 \int_0^1 m^{(1)}\left(x_1 + t b_1 z_1\right) dt, \\
\varphi\left(x_1, \epsilon + m(x_1 + b_1 z_1) + b_1 v_1\right) &= \varphi\left(x_1, \epsilon + m(x_1)\right) + b_1 z_1 \theta_n(x_1, z_1) \int_0^1 \frac{\partial \varphi}{\partial y}\left(x_1, \bar{\theta}_n(u, x_1, z_1)\right) du,
\end{aligned}
$$

where

$$
\theta_n(x_1, z_1) = \int_0^1 m^{(1)}\left(x_1 + t b_1 z_1\right) dt, \quad \bar{\theta}_n(u, x_1, z_1) = \epsilon + m(x_1) + u\theta_n(x_1, z_1).
$$

Therefore

$$
\begin{aligned}
&\int dx_1 \int \left[K_2(v_1)\int \mathbb{1}\left(x_1 + b_1 z_1 \in \mathcal{X}\right) K_1(z_1) dz_1\right]^2 \varphi\left(x_1, \epsilon + m(x_1 + b_1 z_1) + b_1 v_1\right) dv_1 \\
&= \int dx_1 \int \left[K_2(v_1)\int \mathbb{1}\left(x_1 + b_1 z_1 \in \mathcal{X}\right) K_1(z_1) dz_1\right]^2 \varphi\left(x_1, \epsilon + m(x_1)\right) dv_1 + O(b_1) \\
&= \int dx_1 \int \left[\mathbb{1}\left(x_1 \in \mathcal{X}\right) K_2(v_1)\int K_1(z_1) dz_1\right]^2 \varphi\left(x_1, \epsilon + m(x_1)\right) dv_1 \\
&\quad + \int dx_1 \int \delta_n\left(x_1, v_1\right) \varphi\left(x_1, \epsilon + m(x_1)\right) dv_1 + O(b_1), \qquad \text{(B.20)}
\end{aligned}
$$

where

$$\delta_n(x_1, v_1) = \left[K_2(v_1)\int \mathbb{1}\left(x_1 + b_1 z_1 \in \mathcal{X}\right) K_1(z_1) dz_1\right]^2 - \left[\mathbb{1}\left(x_1 \in \mathcal{X}\right) K_2(v_1)\int K_1(z_1) dz_1\right]^2.$$

Applying the Lebesgue Dominated Convergence Theorem yields, for b_1 going to 0,

$$\int dx_1 \int \delta_n\left(x_1, v_1\right) \varphi\left(x_1, \epsilon + m(x_1)\right) dv_1 = o(1).$$

Hence by (B.20) and (4.2.4), we have, since $\int K_1(z_1)dz_1 = 1$ under (H_7),

$$\begin{aligned}
&\int dx_1 \int \left[K_2(v_1)\int \mathbb{1}\left(x_1 + b_1 z_1 \in \mathcal{X}\right) K_1(z_1) dz_1\right]^2 \varphi\left(x_1, \epsilon + m(x_1 + b_1 z_1) + b_1 v_1\right) dv_1 \\
&= \int K_2^2(v_1) dv_1 \int \mathbb{1}\left(x_1 \in \mathcal{X}\right) \varphi\left(x_1, \epsilon + m(x_1)\right) dx_1 + o(1) \\
&= f(\epsilon) \int K_2^2(v) dv + o(1).
\end{aligned}$$

Combining this result with (B.19) and (B.18), we arrive at

$$\operatorname{Var}\left(\widetilde{\widetilde{f}}_{in}(\epsilon)\right) = \frac{f(\epsilon)}{h}\int K_2^2(v)dv + o\left(\frac{1}{h}\right),$$

which proves the second result of the lemma.

The last statement of Lemma is immediate. Indeed, the Triangular and Convex inequalities and the Lebesgue Dominated Convergence Theorem give, by (4.2.4),

$$\begin{aligned}
&\mathbb{E}\left|\widetilde{\widetilde{f}}_{in}(\epsilon) - \mathbb{E}\widetilde{\widetilde{f}}_{in}(\epsilon)\right|^3 \\
&\leq \frac{C}{b_1^{3d}h^3}\int dx_1 \int \left|\int \mathbb{1}\left(x \in \mathcal{X}\right) K_1\left(\frac{x_1 - x}{b_1}\right) K_2\left(\frac{y_1 - \epsilon - m(x)}{b_1}\right) dx\right|^3 dy_1 \\
&= \frac{Cb_1^{3d}h}{b_1^{3d}h^3}\int dx_1 \int \left|\int \mathbb{1}\left(x_1 + b_1 z_1 \in \mathcal{X}\right) K_1(z_1) K_2(v_1) dz_1\right|^3 \varphi\left(x_1, \epsilon + m(x_1 + b_1 z_1) + b_1 v_1\right) dv_1 \\
&= \frac{Cf(\epsilon)}{h^2}\int\int \left|K_1(z_1)K_2(v_1)\right|^3 dz_1 dv_1 + o\left(\frac{1}{h^2}\right). \square
\end{aligned}$$

Appendix C

Proof of Lemma 4.7

See the proof of Lemma 3.1 in Chapter 3. □

Proof of Lemma 4.8

For the first bound in (B.1), set $f_p(e) = e^p f(e)$, and observe if $X_i = x$, we have by (H_4) and the change of variable $e = y - m(x) + hv$,

$$\begin{aligned}\mathbb{E}_{in}\left[\varepsilon_i^p K_2^{(1)}\left(\frac{Y_i - y}{h}\right)\right] &= \mathbb{E}\left[\varepsilon_i^p K_2^{(1)}\left(\frac{\varepsilon_i + m(x) - y}{h}\right)\right] \\ = \quad & \int K_2^{(1)}\left(\frac{e + m(x) - y}{h}\right) f_p(e)de = h\int K_2^{(1)}(v) f_p\left(y - m(x) + hv\right)dv. \qquad (C.1)\end{aligned}$$

Therefore, since f_p has a bounded continuous derivative under (A_5) and that $\int K_2^{(1)}(v)dv = 0$ under (H_8), the Taylor inequality gives

$$\begin{aligned}\left|\int K_2^{(1)}\left(\frac{e + m(x) - y}{h}\right) f_p(e)de\right| &= h\left|\int K_2^{(1)}(v)\left[f_p\left(y - m(x) + hv\right) - f_p\left(y - m(x)\right)\right]\right| dv \\ &\leq h^2 \sup_{u\in\mathbb{R}} |f_p^{(1)}(u)| \int |vK_2^{(1)}(v)|dv \\ &\leq Ch^2,\end{aligned}$$

uniformly in $x \in \mathcal{X}$ and $y \in \mathbb{R}$. Hence from this inequality and (C.1), we deduce

$$\left|\mathbb{E}_{in}\left[\varepsilon_i^p K_2^{(1)}\left(\frac{Y_i - y}{h}\right)\right]\right| \leq Ch^2,$$

for any $y \in \mathbb{R}$. This proves the first inequality in (B.1). The second bound of (B.1) is immediate under (H_5) and (H_8), since for any x in $\mathcal{X}$, $\ell \in [1, 3]$ and $y \in \mathbb{R}$,

$$\begin{aligned}\left|\mathbb{E}_{in}\left[\varepsilon_i^p K_2^{(\ell)}\left(\frac{Y_i - y}{h}\right)^2 \mid X_i = x\right]\right| &= \left|h\int K_2^{(\ell)}(v)^2 f_p\left((y - m(x) + hv\right)dv\right| \\ &\leq h\sup_{u\in\mathbb{R}} |f_p(u)| \int K_2^{(\ell)}(v)^2 dv \\ &\leq Ch, \qquad (C.2)\end{aligned}$$

uniformly for i, x and y. This proves (B.1).

The proof of the second inequalities of (B.2) and (B.3) follows from (C.2). The first bounds in (B.2) and (B.3) are proved simultaneously. For any integer ℓ in $\in \{2, 3\}$ and $x \in \mathcal{X}$, we have

$$\begin{aligned}\mathbb{E}_{in}\left[\varepsilon_i^p K_2^{(\ell)}\left(\frac{Y_i - y}{h}\right) \mid X_i = x\right] &= \int K_2^{(2)}\left(\frac{e + m(x) - y}{h}\right) f_p(e)de \\ &= h\int K_2^{(2)}(v) f_p\left(y - m(x) + hv\right)dv. \qquad (C.3)\end{aligned}$$

Under (H_8), the Kernel function $K_2(\cdot)$ is symmetric, has a compact support and two continuous derivatives, with $\int K_2^{(\ell)}(v)dv = 0$ and $\int vK_2^{(\ell)}(v)dv = 0$. Therefore, since f_p has a bounded continuous

second order derivative by (H_5), the second order Taylor expansion gives, for some $\theta = \theta(y, x, hv)$,

$$\begin{aligned}
& \left| h \int K_2^{(\ell)}(v) f_p \left(y - m(x) + hv\right) dv \right| \\
= & \left| h \int K_2^{(\ell)}(v) \Big[f_p \left(y - m(x) + hv\right) - f_p \left(y - m(x)\right) \Big] dv \right| \\
= & \left| h \int K_2^{(\ell)}(v) \left[hv f_p^{(1)} \left(y - m(x)\right) + \frac{h^2 v^2}{2} f_p^{(2)} \left(y - m(x) + \theta hv\right) \right] dv \right| \\
= & \left| \frac{h^3}{2} \int v^2 K_2^{(\ell)}(v) f_p^{(2)} \left(y - m(x) + \theta hv\right) dv \right| \\
\leq & \; Ch^3.
\end{aligned}$$

Hence from this bound and (C.3), we deduce

$$\left| \mathbb{E}_{in} \left[\varepsilon_i^p K_2^{(\ell)} \left(\frac{Y - y}{h} \right) \right] \right| \leq Ch^3,$$

uniformly for i and y. This ends proof of the Lemma. □

Proof of Lemma 4.9

We have

$$\mathbb{E}_n [S_n] = \int \mathbb{1} \left(x \in \mathcal{X}\right) \mathbb{E}_n \left[S_n (x)\right] dx, \quad \mathbb{E}_n [T_n] = \int \mathbb{1} \left(x \in \mathcal{X}\right) \mathbb{E}_n \left[T_n (x)\right] dx,$$

with

$$\begin{aligned}
\mathbb{E}_n \left[S_n (x)\right] &= \frac{1}{n b_1^d h^2} \sum_{i=1}^n \beta_{in} (x) K_1 \left(\frac{X_i - x}{b_1} \right) \mathbb{E}_n \left[K_2^{(1)} \left(\frac{Y_i - \epsilon - m(x)}{h} \right) \right], \\
\mathbb{E}_n \left[T_n (x)\right] &= \frac{1}{n b_1^d h^3} \sum_{i=1}^n \left(\beta_{in}^2 (x) + \mathbb{E}_n \left[\Sigma_{in}^2 (x)\right] \right) K_1 \left(\frac{X_i - x}{b_1} \right) \mathbb{E}_n \left[K_2^{(2)} \left(\frac{Y_i - \epsilon - m(x)}{h} \right) \right].
\end{aligned}$$

Observe first that under (H_4), Lemma 4.8-(B.1) and Lemma 4.7 give

$$\begin{aligned}
& \sup_{x \in \mathcal{X}} \left| \frac{1}{n b_1^d h^3} \sum_{i=1}^n \mathbb{E}_n \left[\Sigma_{in}^2 (x)\right] K_1 \left(\frac{X_i - x}{b_1} \right) \mathbb{E}_n \left[K_2^{(2)} \left(\frac{Y_i - \epsilon - m(x)}{h} \right) \right] \right| \\
\leq & \; \frac{Ch^3}{h^3} \sup_{x \in \mathcal{X}} \left| \frac{1}{n b_1^d} \sum_{i=1}^n \frac{\sum_{j=1}^n K_0^2 \left(\frac{X_j - x}{b_0} \right)}{\left(n b_0^d \widehat{g}_n (x) \right)^2} K_1 \left(\frac{X_i - x}{b_1} \right) \right| = O_{\mathbb{P}} \left(\frac{1}{n b_0^d} \right),
\end{aligned}$$

and then

$$\left| \int \mathbb{1} \left(x \in \mathcal{X}\right) \frac{1}{n b_1^d h^3} \sum_{i=1}^n \mathbb{E}_n \left[\Sigma_{in}^2 (x)\right] K_1 \left(\frac{X_i - x}{b_1} \right) \mathbb{E}_n \left[K_2^{(2)} \left(\frac{Y_i - \epsilon - m(x)}{h} \right) \right] dx \right| = O_{\mathbb{P}} \left(\frac{1}{n b_0^d} \right).$$

Consider now

$$V_n (p) = \frac{1}{n b_1^d} \int \mathbb{1} \left(x \in \mathcal{X}\right) \sum_{i=1}^n \left| \beta_{in}^p (x) K_1 \left(\frac{X_i - x}{b_1} \right) \right| dx,$$

which is such that, using Lemma 4.8, the equality and the bound above,

$$\left| \mathbb{E}_n [S_n] \right| \leq C V_n (1), \quad \left| \mathbb{E}_n [T_n] \right| \leq C V_n (2) + O_{\mathbb{P}} \left(\frac{1}{n b_0^d} \right).$$

Since Lemma 4.12-(B.4) ensures that $V_n(p) = O_{\mathbb{P}}\left(b_0^{2p}\right)$ for all integer number $p \in [1,6]$, it then follows that

$$\mathbb{E}_n\left[S_n\right] = O_{\mathbb{P}}\left(b_0^2\right), \quad \mathbb{E}_n\left[T_n\right] = O_{\mathbb{P}}\left(b_0^4 + \frac{1}{nb_0^d}\right).$$

This proves the validity of the lemma. □

Proof of Lemma 4.10

Define $e_{in}(x) = \widehat{m}_{in}(x) - m(x)$, which is such that

$$U_n(x) = U_n(x;p) = \frac{1}{nb_1^d h^{p+1}} \sum_{i=1}^{n} e_{in}^p(x)\, K_1\left(\frac{X_i - x}{b_1}\right) K_2^{(p)}\left(\frac{Y_i - \epsilon - m(x)}{h}\right).$$

Let

$$U_n(p) = \int \mathbb{1}\left(x \in \mathcal{X}\right) U_n\left(x\right) dx,$$

so that $S_n = U_n(1)$ and $T_n = U_n(2)$. Observe now that Lemma 4.11 gives

$$\begin{aligned}
&\operatorname{Var}_n\left(U_n(p)\right) \\
&= \operatorname{Var}_n\left(\int \mathbb{1}\left(x \in \mathcal{X}\right) U_n\left(x\right) dx\right) = \int\int \mathbb{1}\left((x_1,x_2) \in \mathcal{X}^2\right) \operatorname{Cov}_n\left(U_n\left(x_1\right), U_n\left(x_2\right)\right) dx_1 dx_2 \\
&= \int\int \mathbb{1}\left((x_1,x_2) \in \mathcal{X}^2, \|x_2 - x_1\| \leq Cb_0 \vee b_1\right) \operatorname{Cov}_n\left(U_n\left(x_1\right), U_n\left(x_2\right)\right) dx_1 dx_2 \\
&\leq \int\int \mathbb{1}\left((x_1,x_2) \in \mathcal{X}^2, \|x_2 - x_1\| \leq Cb_0 \vee b_1\right) \operatorname{Var}_n^{1/2}\left(U_n\left(x_1\right)\right) \operatorname{Var}_n^{1/2}\left(U_n\left(x_2\right)\right) dx_1 dx_2 \\
&\leq \frac{1}{2}\int\int \mathbb{1}\left((x_1,x_2) \in \mathcal{X}^2, \|x_2 - x_1\| \leq Cb_0 \vee b_1\right) \left\{\operatorname{Var}_n\left(U_n\left(x_1\right)\right) + \operatorname{Var}_n\left(U_n\left(x_2\right)\right)\right\} dx_1 dx_2 \\
&\leq C\left(b_0^d \vee b_1^d\right) \int \mathbb{1}\left(x \in \mathcal{X}\right) \operatorname{Var}_n\left(U_n\left(x\right)\right) dx, \qquad \text{(C.4)}
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
&\left(nb_1^d h^{p+1}\right)^2 \int \mathbb{1}\left(x \in \mathcal{X}\right) \operatorname{Var}_n\left(U_n\left(x\right)\right) dx \\
&= \sum_{i=1}^{n} \int \mathbb{1}\left(x \in \mathcal{X}\right) K_1^2\left(\frac{X_i - x}{b_1}\right) \operatorname{Var}_n\left(W_{in}(x;p)\right) dx \\
&\quad + \sum_{1 \leq i_1 \neq i_2 \leq n} \int \mathbb{1}\left(x \in \mathcal{X}\right) K_1\left(\frac{X_{i_1} - x}{b_1}\right) K_1\left(\frac{X_{i_2} - x}{b_1}\right) \operatorname{Cov}_n\left(W_{i_1 n}(x;p), W_{i_2 n}(x;p)\right) dx, \qquad \text{(C.5)}
\end{aligned}$$

where

$$W_{in}(x;p) = e_{in}^p\left(x\right) K_2^{(p)}\left(\frac{Y_i - \epsilon - m(x)}{h}\right).$$

The first term in (C.5) yields, by Lemma 4.8 and Lemma 4.12,

$$\begin{aligned}
&\sum_{i=1}^{n}\int \mathbb{1}\left(x\in\mathcal{X}\right)K_1^2\left(\frac{X_i-x}{b_1}\right)\mathrm{Var}_n\left(W_{in}(x;p)\right)dx\\
&\leq \sum_{i=1}^{n}\int \mathbb{1}\left(x\in\mathcal{X}\right)K_1^2\left(\frac{X_i-x}{b_1}\right)\mathbb{E}_n\left[e_{in}^{2p}\left(x\right)K_2^{(p)}\left(\frac{Y_i-\epsilon-m(x)}{h}\right)^2\right]dx\\
&= \sum_{i=1}^{n}\int \mathbb{1}\left(x\in\mathcal{X}\right)K_1^2\left(\frac{X_i-x}{b_1}\right)\mathbb{E}_n\left[e_{in}^{2p}\left(x\right)\mathbb{E}_{in}\left[K_2^{(p)}\left(\frac{Y_i-\epsilon-m(x)}{h}\right)^2\right]\right]dx\\
&\leq Ch\sum_{i=1}^{n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\mathbb{E}_n\left[e_{in}^{2p}\left(x\right)\right]K_1^2\left(\frac{X_i-x}{b_1}\right)dx\\
&\leq Ch\sum_{i=1}^{n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\left(\beta_{in}^{2p}\left(x\right)+\mathbb{E}_n\left[\Sigma_{in}^{2p}(x)\right]\right)K_1^2\left(\frac{X_i-x}{b_1}\right)dx\\
&= O_{\mathbb{P}}\left(nb_1^dh\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^p. \qquad \text{(C.6)}
\end{aligned}$$

For the sum of the conditional covariances in (C.5), set

$$\widetilde{W}_n(p)=\sum_{1\leq i_1\neq i_2\leq n}\int \mathbb{1}\left(x\in\mathcal{X}\right)K_1\left(\frac{X_{i_1}-x}{b_1}\right)K_1\left(\frac{X_{i_2}-x}{b_1}\right)\mathrm{Cov}_n\left(W_{i_1n}(x;p),W_{i_2n}(x;p)\right)dx.$$

We need to bound this term for $p\in[1,2]$. Since

$$W_{in}(x;p)=\left(\beta_{in}(x)+\Sigma_{in}(x)\right)^p\left(x\right)K_2^{(p)}\left(\frac{Y_i-\epsilon-m(x)}{h}\right),$$

the independence of the the Y_j's gives, for any $i_1\neq i_2$,

$$\begin{aligned}
&\mathrm{Cov}_n\left(W_{i_1n}(x;1),W_{i_2n}(x;1)\right)\\
&= \beta_{i_1n}(x)\mathrm{Cov}_n\left[K_2^{(1)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right),\Sigma_{i_2n}(x)K_2^{(1)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\\
&\quad+\beta_{i_2n}(x)\mathrm{Cov}_n\left[\Sigma_{i_1n}(x)K_2^{(1)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right),K_2^{(1)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\\
&\quad+\mathrm{Cov}_n\left[\Sigma_{i_1n}(x)K_2^{(1)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right),\Sigma_{i_2n}(x)K_2^{(1)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]. \qquad \text{(C.7)}
\end{aligned}$$

Moreover, it is clear that the results of Lemma 4.8 remain valid with $\mathbb{E}_n[\cdot]$, since $\mathbb{E}_n[A]=\mathbb{E}_n[\mathbb{E}_{in}[A]]$, where $\mathbb{E}_{in}[\cdot]$ represents the conditional mean given $(X_1,\dots,X_n,\varepsilon_k,k\neq i)$. Therefore, since $K_0(\cdot)$ is bounded under (H_7), this yields, by (H_4) and Lemma 4.7,

$$\begin{aligned}
&\left|\beta_{i_1n}(x)\mathrm{Cov}_n\left[K_2^{(1)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right),\Sigma_{i_2n}(x)K_2^{(1)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\right|\\
&= \left|\frac{\beta_{i_1n}(x)}{nb_0^d\widehat{g}_n(x)}K_0\left(\frac{X_{i_2}-x}{b_0}\right)\mathbb{E}_n\left[\varepsilon_{i_1}K_2^{(1)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right)\right]\mathbb{E}_n\left[K_2^{(1)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\right|.\\
&= O_{\mathbb{P}}\left(\frac{h^4}{nb_0^d}\right)\left|\beta_{i_1n}(x)\right|, \qquad \text{(C.8)}
\end{aligned}$$

uniformly in x, i_1 and i_2. We also have

$$\begin{aligned}
&\mathrm{Cov}_n\left[\Sigma_{i_1n}(x)K_2^{(1)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right),\Sigma_{i_2n}(x)K_2^{(1)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\\
&= \frac{\mathbb{E}\left[\varepsilon^2\right]}{(nb_0^d\widehat{g}_n(x))^2}\sum_{\substack{i_3=1\\ i_3\neq i_1,i_2}}^{n} K_0^2\left(\frac{X_{i_3}-x}{b_0}\right)\mathbb{E}_n\left[K_2^{(1)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right)\right]\mathbb{E}_n\left[K_2^{(1)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\\
&\quad+\frac{1}{(nb_0^d\widehat{g}_n(x))^2}K_0\left(\frac{X_{i_1}-x}{b_0}\right)\mathbb{E}_n\left[\varepsilon_{i_1}K_2^{(1)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right)\right]\\
&\qquad\times K_0\left(\frac{X_{i_2}-x}{b_0}\right)\mathbb{E}_n\left[\varepsilon_{i_2}K_2^{(1)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right].
\end{aligned}$$

This gives, by (H_4), (H_7), Lemma 4.7 and Lemma 4.8,

$$\begin{aligned}
&\left|\mathrm{Cov}_n\left[\Sigma_{i_1n}(x)K_2^{(1)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right),\Sigma_{i_2n}(x)K_2^{(1)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\right|\\
&= O_{\mathbb{P}}\left(\frac{h^4}{(nb_0^d)^2}\right)\sum_{i_3=1}^{n}K^2\left(\frac{X_{i_3}-x}{b_0}\right)+O_{\mathbb{P}}\left(\frac{h^4}{(nb_0^d)^2}\right)=O_{\mathbb{P}}\left(\frac{h^4}{nb_0^d}\right),
\end{aligned}$$

uniformly for any $i_1 \neq i_2$, x_1 and x_2. Collecting this result, (C.8) and (C.7), it follows, using Lemma 4.7 and taking $p_1 = 1$ in Lemma 4.12-(B.4),

$$\begin{aligned}
&O_{\mathbb{P}}\left(\frac{nb_0^d}{h^4}\right)\widetilde{W}_n(1)\\
&= \sum_{1\le i_1\neq i_2\le n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\left|\beta_{i_1n}(x)K_1\left(\frac{X_{i_1}-x}{b_1}\right)K_1\left(\frac{X_{i_2}-x}{b_1}\right)\right|dx\\
&\quad+\sum_{1\le i_1\neq i_2\le n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\left|K_1\left(\frac{X_{i_1}-x}{b_1}\right)K_1\left(\frac{X_{i_2}-x}{b_1}\right)\right|dx\\
&\le O_{\mathbb{P}}\left(nb_1^d\right)\sum_{i=1}^{n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\left|\beta_{in}(x)K_1\left(\frac{X_i-x}{b_1}\right)\right|dx+O_{\mathbb{P}}\left(n^2b_1^{2d}\right)\\
&= O_{\mathbb{P}}\left(n^2b_1^{2d}\right)\left(b_0^2\right)+O_{\mathbb{P}}\left(n^2b_1^{2d}\right)=O_{\mathbb{P}}\left(n^2b_1^{2d}\right).
\end{aligned}$$

Combining this result with (C.6), (C.5) and (C.4), we arrive at

$$\begin{aligned}
&\mathrm{Var}_n\left(S_n\right)=\mathrm{Var}_n\left(U_n(1)\right)\\
&\le O_{\mathbb{P}}\left(b_0^d\vee b_1^d\right)\times\frac{1}{\left(nb_1^dh^2\right)^2}\left[nb_1^dh\left(b_0^4+\frac{1}{nb_0^d}\right)+\widetilde{W}_n(1)\right]\\
&= O_{\mathbb{P}}\left(b_0^d\vee b_1^d\right)\times\frac{1}{\left(nb_1^dh^2\right)^2}\left[nb_1^dh\left(b_0^4+\frac{1}{nb_0^d}\right)+\frac{nb_1^{2d}h^4}{b_0^d}\right]\\
&= O_{\mathbb{P}}\left(b_0^d\vee b_1^d\right)\left[\frac{1}{nb_1^dh^3}\left(b_0^4+\frac{1}{nb_0^d}\right)+\frac{1}{nb_0^d}\right].
\end{aligned}$$

This proves the first result of the Lemma.

For the second, we also have by (C.4), (C.5) and (C.6),

$$\mathrm{Var}_n\left(T_n\right)=\mathrm{Var}_n\left(U_n(2)\right)=\frac{O_{\mathbb{P}}\left(b_0^d\vee b_1^d\right)}{\left(nb_1^dh^3\right)^2}\left[nb_1^dh\left(b_0^4+\frac{1}{nb_0^d}\right)^2+\widetilde{W}_n(2)\right].$$

Hence the order of $\mathrm{Var}_n\left(T_n\right)$ follows from the following result

$$\widetilde{W}_n(2)=O_{\mathbb{P}}\left[\frac{h^6}{nb_0^d}\left(n^2b_1^{2d}\right)\left(b_0^4\right)+\frac{h^3}{n^2b_0^{2d}}\left(n^2b_1^{2d}\right)\right]. \tag{C.9}$$

Indeed, (C.9) and the equality before give

$$
\begin{aligned}
&\mathrm{Var}_n\left(T_n\right)\\
&= \frac{O_{\mathbb{P}}\left(b_0^d \vee b_1^d\right)}{\left(nb_1^d h^3\right)^2}\left[nb_1^d h\left(b_0^4+\frac{1}{nb_0^d}\right)^2+\frac{h^6}{nb_0^d}\left(n^2b_1^{2d}\right)\left(b_0^4\right)+\frac{h^3}{n^2b_0^{2d}}\left(n^2b_1^{2d}\right)\right]\\
&= O_{\mathbb{P}}\left(b_0^d \vee b_1^d\right)\left[\frac{1}{nb_1^d h^5}\left(b_0^4+\frac{1}{nb_0^d}\right)^2+\frac{b_0^4}{nb_0^d}+\frac{1}{n^2b_0^{2d}h^3}\right].
\end{aligned}
$$

This yields the second result of the Lemma. We now prove (C.9). Observe that for $i_1 \neq i_2$, we have

$$
\begin{aligned}
&\mathrm{Cov}_n\left(W_{i_1n}(x;2), W_{i_2n}(x;2)\right)\\
&= \mathrm{Cov}_n\left[\left(\beta_{i_1n}(x)+\Sigma_{i_1n}(x)\right)^2 K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right), \left(\beta_{i_2n}(x)+\Sigma_{i_2n}(x)\right)^2 K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\\
&= \beta_{i_1n}^2(x)\mathrm{Cov}_n\left[K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right), \Sigma_{i_2n}^2(x)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\\
&\quad+\beta_{i_2n}^2(x)\mathrm{Cov}_n\left[\Sigma_{i_1n}^2(x)K_2^{(1)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right), K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\\
&\quad+2\beta_{i_1n}^2(x)\beta_{i_2n}(x)\mathrm{Cov}_n\left[K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right), \Sigma_{i_2n}(x)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\\
&\quad+2\beta_{i_1n}(x)\beta_{i_2n}^2(x)\mathrm{Cov}_n\left[\Sigma_{i_1n}(x)K_2^{(1)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right), K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\\
&\quad+2\beta_{i_1n}(x)\mathrm{Cov}_n\left[\Sigma_{i_1n}(x)K_2^{(1)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right), \Sigma_{i_2n}^2(x)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\\
&\quad+2\beta_{i_2n}(x)\mathrm{Cov}_n\left[\Sigma_{i_1n}^2(x)K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right), \Sigma_{i_1n}(x)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\\
&\quad+4\beta_{i_1n}(x)\beta_{i_2n}(x)\mathrm{Cov}_n\left[\Sigma_{i_1n}(x)K_2^{(1)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right), \Sigma_{i_2n}(x)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\\
&\quad+\mathrm{Cov}_n\left[\Sigma_{i_1n}^2(x)K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right), \Sigma_{i_2n}^2(x)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]. \qquad \text{(C.10)}
\end{aligned}
$$

The two-first terms in (C.10) are treated similarly, since they are symmetric. Under (H_4), we have, for any $i_1 \neq i_2$,

$$
\begin{aligned}
&\beta_{i_1n}^2(x)\mathrm{Cov}_n\left[K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right), \Sigma_{i_2n}^2(x)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\\
&= \frac{\beta_{i_1n}^2(x)}{\left(nb_0^d\widehat{g}_n(x)\right)^2}\sum_{1\leq i_3\neq i_2\leq n} K_0^2\left(\frac{X_{i_3}-x}{b_0}\right)\mathrm{Cov}_n\left[K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right), \varepsilon_{i_3}^2 K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\\
&= \frac{\beta_{i_1n}^2(x)}{\left(nb_0^d\widehat{g}_n(x)\right)^2}K_0^2\left(\frac{X_{i_1}-x}{b_0}\right)\mathrm{Cov}_n\left[K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right), \varepsilon_{i_1}^2 K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right],
\end{aligned}
$$

with, using Lemma 4.8,

$$
\begin{aligned}
&\left|\mathrm{Cov}_n\left[K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right), \varepsilon_{i_1}^2 K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\right|\\
&\leq \left|\mathbb{E}_n\left[\varepsilon_{i_1}^2 K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\right|\\
&\quad+\left|\mathbb{E}_n\left[\varepsilon_{i_1}^2 K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\mathbb{E}_n\left[K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right)\right]\right|\\
&\leq Ch^6.
\end{aligned}
$$

Therefore, since $K_0(\cdot)$ is bounded under (H_7), Lemma 4.7 gives

$$\begin{aligned}
&\left|\beta^2_{i_1n}(x)\mathrm{Cov}_n\left[K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right),\Sigma^2_{i_2n}(x)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\right| \\
&\leq \frac{Ch^6\beta^2_{i_1n}(x)}{\left(nb_0^d\widehat{g}_n(x)\right)^2}K_0^2\left(\frac{X_{i_1}-x}{b_0}\right)=O_{\mathbb{P}}\left(\frac{h^6}{(nb_0^d)^2}\right)\beta^2_{i_1n}(x), \qquad \text{(C.11)}
\end{aligned}$$

uniformly with respect to i_1, i_2 and x.

For the third and the fourth in (C.10), we also have, uniformly for i_1, i_2 and x,

$$\begin{aligned}
&\left|\beta_{i_1n}(x)\beta^2_{i_2n}(x)\mathrm{Cov}_n\left[\Sigma_{i_1n}(x)K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right),K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\right| \\
&= \left|\frac{\beta_{i_1n}(x)\beta^2_{i_2n}(x)}{nb_0^d\widehat{g}_n(x)}K_0\left(\frac{X_{i_1}-x}{b_0}\right)\mathbb{E}_n\left[\varepsilon_{i_2}K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right)\right]\mathbb{E}_n\left[K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right)\right]\right| \\
&= O_{\mathbb{P}}\left(\frac{h^6}{nb_0^d}\right)\left|\beta_{i_1n}(x)\beta^2_{i_2n}(x)\right|. \qquad \text{(C.12)}
\end{aligned}$$

Further, note that

$$\begin{aligned}
&\beta_{i_1n}(x)\mathrm{Cov}_n\left[\Sigma_{i_1n}(x)K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right),\Sigma^2_{i_2n}(x)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right] \\
&= \beta_{i_1n}(x)\mathbb{E}_n\left[\Sigma_{i_1n}(x)\Sigma^2_{i_2n}(x)K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right] \\
&= \beta_{i_1n}(x)\mathbb{E}_n\left[\Sigma^2_{i_2n}(x)K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right)\mathbb{E}_{i_2n}\left[\Sigma_{i_1n}(x)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\right],
\end{aligned}$$

where

$$\begin{aligned}
&\left|\mathbb{E}_{i_2n}\left[\Sigma_{i_1n}(x)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\right| \\
&= \left|\frac{1}{nb_0^d\widehat{g}_n(x)}K\left(\frac{X_{i_2}-x}{b_0}\right)\mathbb{E}_n\left[\varepsilon_{i_2}K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\right| \leq \frac{Ch^3}{nb_0^d\left|\widehat{g}_n(x)\right|}.
\end{aligned}$$

Therefore by (H_7) and Lemma 4.7, we have, uniformly for i_1, i_2 and x,

$$\begin{aligned}
&\left|\beta_{i_1n}(x)\mathrm{Cov}_n\left[\Sigma_{i_1n}(x)K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right),\Sigma^2_{i_2n}(x)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\right| \\
&= O_{\mathbb{P}}\left(\frac{h^3}{nb_0^d}\right)\left|\beta_{i_1n}(x)\right|\mathbb{E}_n\left[\Sigma^2_{i_2n}(x)\right] \\
&\leq O_{\mathbb{P}}\left(\frac{h^3}{nb_0^d}\right)\left|\beta_{i_1n}(x)\right|\times\frac{\mathbb{E}\left[\varepsilon^2\right]}{\left(nb_0^d\widehat{g}_n(x)\right)^2}\sum_{j=1}^n K_0^2\left(\frac{X_j-x}{b_0}\right) \\
&\leq O_{\mathbb{P}}\left(\frac{h^3}{n^2b_0^{2d}}\right)\left|\beta_{i_1n}(x)\right|, \qquad \text{(C.13)}
\end{aligned}$$

We now treat the two last terms in (C.10). Observe that

$$\begin{aligned}
&\mathrm{Cov}_n\left[\Sigma_{i_1n}(x)K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right),\Sigma_{i_2n}(x)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right] \\
&= \frac{\mathbb{E}\left[\varepsilon^2\right]}{(nb_0^d\widehat{g}_n(x))^2}\sum_{\substack{i_3=1\\ i_3\neq i_1,i_2}}^n K_0^2\left(\frac{X_{i_3}-x}{b_0}\right)\mathbb{E}_n\left[K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right)\right]\mathbb{E}_n\left[K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right] \\
&\quad+\frac{1}{(nb_0^d\widehat{g}_n(x))^2}K_0\left(\frac{X_{i_1}-x}{b_0}\right)\mathbb{E}_n\left[\varepsilon_{i_1}K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right)\right] \\
&\qquad\times K_0\left(\frac{X_{i_2}-x}{b_0}\right)\mathbb{E}_n\left[\varepsilon_{i_2}K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right].
\end{aligned}$$

This gives, by Lemma 4.7, Lemma 4.8 and uniformly with respect to $i_1 \neq i_2$ and x,

$$\begin{aligned}
&\left|\beta_{i_1 n}(x)\beta_{i_2 n}(x)\mathrm{Cov}_n\left[\Sigma_{i_1 n}(x)K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right), \Sigma_{i_2 n}(x)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\right| \\
&= O_{\mathbb{P}}\left[\frac{h^6}{(nb_0^d)^2}\sum_{i_3=1}^{n}K_0^2\left(\frac{X_{i_3}-x}{b_0}\right)+\frac{h^6}{(nb_0^d)^2}\right]|\beta_{i_1 n}(x)\beta_{i_2 n}(x)| \\
&= O_{\mathbb{P}}\left(\frac{h^6}{nb_0^d}\right)|\beta_{i_1 n}(x)\beta_{i_2 n}(x)|. \qquad \text{(C.14)}
\end{aligned}$$

Moreover,

$$\begin{aligned}
&\left|\mathrm{Cov}_n\left[\Sigma_{i_1 n}^2(x)K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right), \Sigma_{i_2 n}^2(x)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\right| \\
&\leq \left|\mathbb{E}_n\left[\Sigma_{i_1 n}^2(x)\Sigma_{i_2 n}^2(x)K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\right| \\
&\quad +\left|\mathbb{E}_n\left[\Sigma_{i_1 n}^2(x)K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right)\right]\mathbb{E}_n\left[\Sigma_{i_2 n}^2(x)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\right|, \qquad \text{(C.15)}
\end{aligned}$$

with, using (H_4), Lemma 4.8 and Lemma 4.7,

$$\begin{aligned}
\left|\mathbb{E}_n\left[\Sigma_{in}^2(x)K_2^{(2)}\left(\frac{Y_i-\epsilon-m(x)}{h}\right)\right]\right| &= \left|\mathbb{E}_n\left[\Sigma_{in}^2(x)\right]\mathbb{E}_n\left[K_2^{(2)}\left(\frac{Y_i-\epsilon-m(x)}{h}\right)\right]\right| \\
&\leq \frac{Ch^3}{\left(nb_0^d\widehat{g}_n(x)\right)^2}\sum_{j=1}^{n}K_0^2\left(\frac{X_j-x}{b_0}\right) \\
&= O_{\mathbb{P}}\left(\frac{h^3}{nb_0^d}\right), \qquad \text{(C.16)}
\end{aligned}$$

uniformly for i and x. Moreover, for the first term in Bound (C.10), we have

$$\begin{aligned}
&\mathbb{E}_n\left[\Sigma_{i_1 n}^2(x)\Sigma_{i_2 n}^2(x)K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right] \\
&= \mathbb{E}_n\left[\Sigma_{i_1 n}^2(x)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\mathbb{E}_{i_1 n}\left[\Sigma_{i_2 n}^2(x)K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right)\right]\right],
\end{aligned}$$

where

$$\begin{aligned}
&\left|\mathbb{E}_{i_1 n}\left[\Sigma_{i_2 n}^2(x)K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right)\right]\right| \\
&= \left|\frac{1}{\left(nb_0^d\widehat{g}_n(x)\right)^2}\sum_{1\leq i_3\neq i_2\leq n}K_0^2\left(\frac{X_{i_3}-x}{b_0}\right)\mathbb{E}_{i_1 n}\left[\varepsilon_{i_3}K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right)\right]\right| \\
&\leq \frac{Ch^3}{nb_0^d\left(\widehat{g}_n(x)\right)^2}.
\end{aligned}$$

Therefore, since $K_2^{(2)}$ is bounded under (H_7), it follows, by Lemma 4.7, and uniformly with respect to x, i_1 and i_2,

$$\begin{aligned}
&\left|\mathbb{E}_n\left[\Sigma_{i_1 n}^2(x)\Sigma_{i_2 n}^2(x)K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\right| \\
&= O_{\mathbb{P}}\left(\frac{h^3}{nb_0^d}\right)\mathbb{E}_n\left[\Sigma_{i_2 n}^2(x)\right] = O_{\mathbb{P}}\left(\frac{h^3}{n^2b_0^{2d}}\right),
\end{aligned}$$

Hence from (C.16) and (C.15), we deduce

$$\left|\mathrm{Cov}_n\left[\Sigma_{i_1n}^2(x)K_2^{(2)}\left(\frac{Y_{i_1}-\epsilon-m(x)}{h}\right),\Sigma_{i_2n}^2(x)K_2^{(2)}\left(\frac{Y_{i_2}-\epsilon-m(x)}{h}\right)\right]\right| = O_{\mathbb{P}}\left(\frac{h^3}{n^2b_0^{2d}}\right),$$

uniformly in x, i_1 and i_2. Collecting this result, (C.13)-(C.14) and (C.11)-(C.12), it follows then by Equality (C.10),

$$\begin{aligned}
\left|\widetilde{W}_n(2)\right| &= \left|\sum_{1\le i_1\neq i_2\le n}\int \mathbb{1}\left(x\in\mathcal{X}\right)K_1\left(\frac{X_{i_1}-x}{b_1}\right)K_1\left(\frac{X_{i_2}-x}{b_1}\right)\mathrm{Cov}_n\left(W_{i_1n}(x;2),W_{i_2n}(x;2)\right)dx\right| \\
&= O_{\mathbb{P}}\left(\frac{h^6}{(nb_0^d)^2}\right)\sum_{1\le i_1\neq i_2\le n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\left|\beta_{i_1n}^2(x)K_1\left(\frac{X_{i_1}-x}{b_1}\right)K_1\left(\frac{X_{i_2}-x}{b_1}\right)\right|dx \\
&\quad +O_{\mathbb{P}}\left(\frac{h^6}{nb_0^d}\right)\sum_{1\le i_1\neq i_2\le n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\left|\beta_{i_1n}(x)\beta_{i_2n}^2(x)K_1\left(\frac{X_{i_1}-x}{b_1}\right)K_1\left(\frac{X_{i_2}-x}{b_1}\right)\right|dx \\
&\quad +O_{\mathbb{P}}\left(\frac{h^3}{n^2b_0^{2d}}\right)\sum_{1\le i_1\neq i_2\le n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\left|\beta_{i_1n}(x)K_1\left(\frac{X_{i_1}-x}{b_1}\right)K_1\left(\frac{X_{i_2}-x}{b_1}\right)\right|dx \\
&\quad +O_{\mathbb{P}}\left(\frac{h^6}{nb_0^d}\right)\sum_{1\le i_1\neq i_2\le n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\left|\beta_{i_1n}(x)\beta_{i_2n}(x)K_1\left(\frac{X_{i_1}-x}{b_1}\right)K_1\left(\frac{X_{i_2}-x}{b_1}\right)\right|dx \\
&\quad +O_{\mathbb{P}}\left(\frac{h^3}{n^2b_0^{2d}}\right)\sum_{1\le i_1\neq i_2\le n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\left|K_1\left(\frac{X_{i_1}-x}{b_1}\right)K_1\left(\frac{X_{i_2}-x}{b_1}\right)\right|dx. \qquad \text{(C.17)}
\end{aligned}$$

Moreover, note that for any integers p_1 and p_2 in $[0,2]$,

$$\begin{aligned}
&\sum_{1\le i_1\neq i_2\le n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\left|\beta_{i_1n}^{p_1}(x)\beta_{i_2n}^{p_2}(x)K_1\left(\frac{X_{i_1}-x}{b_1}\right)K_1\left(\frac{X_{i_2}-x}{b_1}\right)\right|dx \\
&\le \sum_{1\le i_1\neq i_2\le n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\left|\beta_{i_1n}^{p_1+p_2}(x)K_1\left(\frac{X_{i_1}-x}{b_1}\right)K_1\left(\frac{X_{i_2}-x}{b_1}\right)\right|dx \\
&\quad + \sum_{1\le i_1\neq i_2\le n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\left|\beta_{i_2n}^{p_1+p_2}(x)K_1\left(\frac{X_{i_1}-x}{b_1}\right)K_1\left(\frac{X_{i_2}-x}{b_1}\right)\right|dx.
\end{aligned}$$

Since (H_7) and Lemma 4.12-(B.4) give, for $p=p_1+p_2$,

$$\begin{aligned}
&\sum_{1\le i_1\neq i_2\le n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\left|\beta_{i_1n}^{p_1+p_2}(x)K_1\left(\frac{X_{i_1}-x}{b_1}\right)K_1\left(\frac{X_{i_2}-x}{b_1}\right)\right|dx \\
&= b_1^d\sum_{1\le i_1\neq i_2\le n}\int \mathbb{1}\left(u+b_1X_{i_2}\in\mathcal{X}\right)\left|\beta_{i_1n}^{p_1+p_2}\left(u+b_1X_{i_2}\right)K_1\left(u\right)K_1\left(\frac{X_{i_2}-u-b_1X_{i_2}}{b_1}\right)\right|du \\
&= O_{\mathbb{P}}\left(nb_1^d\right)\sum_{i=1}^{n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\left|\beta_{in}^{p_1+p_2}(x)K_1\left(\frac{X_i-x}{b_1}\right)\right|dx \\
&= O_{\mathbb{P}}\left(n^2b_1^{2d}\right)\left(b_0^{2p}\right),
\end{aligned}$$

it the follows, by this result, the bound above and (C.17),

$$\widetilde{W}_n(2) = O_{\mathbb{P}}\left[\frac{h^6}{nb_0^d}\left(n^2b_1^{2d}\right)\left(b_0^4\right)+\frac{h^3}{n^2b_0^{2d}}\left(n^2b_1^{2d}\right)\right].$$

This proves (C.9) and then completes the proof of the Lemma. □

Proof of Lemma 4.11

The lemma follows directly from the fact that given $X_1, \ldots, X_n$, we have $U_n(x_1) = \Phi_{1n}(\varepsilon_i, i \in I_1)$ and $U_n(x_2) = \Phi_{2n}(\varepsilon_i, i \in I_2)$, with an empty $I_1 \cap I_2$, since the Kernel functions are compactly supported and $\|x_2 - x_1\| \geq Cb_0 \vee b_1$ for a sufficiently large C. $\square$

Proof of Lemma 4.12

Define

$$V_n = \frac{1}{nb_1^d} \int \mathbb{1}(x \in \mathcal{X}) \sum_{i=1}^{n} \left| \beta_{in}^{p_1}(x) K_1^{p_2}\left(\frac{X_i - x}{b_1}\right)\right| dx,$$

and

$$\Delta_j(x) = (m(X_j) - m(x)) K_0\left(\frac{X_j - x}{b_0}\right),$$

which is such that, using Lemma 4.7,

$$\begin{aligned}
|\beta_{in}(x)| &= \left| \frac{\sum_{1 \leq j \neq i \leq n} \Delta_j(x)}{nb_0^d \widehat{g}_n(x)} \right| \\
&\leq \sup_{x \in \mathcal{X}} \left| \frac{1}{\widehat{g}_n(x)} \right| \times \frac{1}{nb_0^d} \left(\left| \sum_{1 \leq j \neq i \leq n} (\Delta_j(x) - \mathbb{E}[\Delta_j(x)]) \right| + \left| \sum_{1 \leq j \neq i \leq n} \mathbb{E}[\Delta_j(x)] \right| \right) \\
&\leq \frac{O_{\mathbb{P}}(1)}{nb_0^d} \left(\left| \sum_{1 \leq j \neq i \leq n} (\Delta_j(x) - \mathbb{E}[\Delta_j(x)]) \right| + \left| \sum_{1 \leq j \neq i \leq n} \mathbb{E}[\Delta_j(x)] \right| \right),
\end{aligned}$$

uniformly in x. This gives using the Markov Inequality which ensures that $A_n = O_{\mathbb{P}}(\mathbb{E}|A_n|)$,

$$\begin{aligned}
&|V_n| \\
&\leq \frac{O_{\mathbb{P}}(1)}{nb_1^d} \frac{1}{(nb_0^d)^{p_1}} \sum_{i=1}^{n} \int \mathbb{1}(x \in \mathcal{X}) \left\{ \left| \sum_{1 \leq j \neq i \leq n} (\Delta_j(x) - \mathbb{E}[\Delta_j(x)]) \right| + \left| \sum_{1 \leq j \neq i \leq n} \mathbb{E}[\Delta_j(x)] \right| \right\}^{p_1} \\
&\qquad \times \mathbb{E}\left| K_1^{p_2}\left(\frac{X_i - x}{b_1}\right)\right| dx \\
&\leq \frac{O_{\mathbb{P}}(1)}{(nb_0^d)^{p_1}} \int \mathbb{1}(x \in \mathcal{X}) \left\{ \mathbb{E}\left[\left| \sum_{j=2}^{n} (\Delta_j(x) - \mathbb{E}[\Delta_j(x)]) \right|^{p_1} \right] + \left| \sum_{j=2}^{n} \mathbb{E}[\Delta_j(x)] \right|^{p_1} \right\} dx. \qquad \text{(C.18)}
\end{aligned}$$

We bound the two resulting integrals in (C.18). For the first, the Marcinkiewicz-Zygmund inequality (see e.g Chow and Teicher, 2003, p. 386), the Hölder and the Minkowski inequalities give

$$\begin{aligned}
&\int \mathbb{1}(x\in\mathcal{X})\,\mathbb{E}\left[\left|\sum_{j=2}^{n}(\Delta_j(x)-\mathbb{E}[\Delta_j(x)])\right|^{p_1}\right]dx\\
\leq\;& \int \mathbb{1}(x\in\mathcal{X})\,\mathbb{E}^{1/2}\left[\left|\sum_{j=2}^{n}(\Delta_j(x)-\mathbb{E}[\Delta_j(x)])\right|^{2p_1}\right]dx\\
\leq\;& \int \mathbb{1}(x\in\mathcal{X})\,\mathbb{E}^{1/2}\left[\left|\sum_{j=2}^{n}(\Delta_j(x)-\mathbb{E}[\Delta_j(x)])^2\right|^{p_1}\right]dx\\
=\;& \int \mathbb{1}(x\in\mathcal{X})\left\{\mathbb{E}^{1/p_1}\left[\left|\sum_{j=2}^{n}(\Delta_j(x)-\mathbb{E}[\Delta_j(x)])^2\right|^{p}\right]\right\}^{p_1/2}dx\\
\leq\;& \int \mathbb{1}(x\in\mathcal{X})\left\{\sum_{j=2}^{n}\mathbb{E}^{1/p_1}\left[\left|(\Delta_j(x)-\mathbb{E}[\Delta_j(x)])^2\right|^{p_1}\right]\right\}^{p_1/2}dx\\
\leq\;& C\int \mathbb{1}(x\in\mathcal{X})\left\{\sum_{j=2}^{n}\mathbb{E}^{1/p_1}\left[\Delta_j^{2p_1}(x)\right]\right\}^{p_1/2}dx\\
=\;& C\int \mathbb{1}(x\in\mathcal{X})\left\{\sum_{j=2}^{n}\left[\int\left((m(z)-m(x))K_0\left(\frac{z-x}{b_0}\right)\right)^{2p_1}g(z)\,dz\right]^{1/p_1}\right\}^{p_1/2}dx\\
=\;& C\int \mathbb{1}(x\in\mathcal{X})\left\{\sum_{j=2}^{n}\left[b_0^d\int((m(x+b_0u)-m(x))K_0(u))^{2p}g(x+b_0u)\,du\right]^{1/p_1}\right\}^{p_1/2}dx\\
\leq\;& C\left\{n\left[b_0^d b_0^{2p_1}\right]^{1/p_1}\right\}^{p_1/2}=O\left(\left(n^{p_1}b_0^d\right)^{1/2}b_0^{p_1}\right).
\end{aligned}\tag{C.19}$$

For the second resulting integral in (C.18), we have, since the $\Delta_j(x)$'s are identically distributed,

$$\begin{aligned}
&\int \mathbb{1}(x\in\mathcal{X})\left|\sum_{j=2}^{n}\mathbb{E}[\Delta_j(x)]\right|^{p_1}dx \leq n^{p_1}\int \mathbb{1}(x\in\mathcal{X})\left|\mathbb{E}[\Delta_1(x)]\right|^{p_1}dx\\
\leq\;& n^{p_1}\int \mathbb{1}(x\in\mathcal{X})\left|b_0^d\int(m(x+b_0u)-m(x))\,g(x+b_0u)\,K_0(u)\,du\right|^{p_1}dx\\
\leq\;& Cn^{p_1}\left[\left(b_0^d\times b_0^2\right)^{p_1}\right]=O\left(nb_0^{d+2}\right)^{p_1},
\end{aligned}$$

using $\int uK_0(u)du=0$ and the fact that expect for those x at a distance $O(b_0)$ of the boundaries of $\mathcal{X}$, we have for all u in the support of $K_0(\cdot)$,

$$\begin{aligned}
&(m(x+b_0u)-m(x))\,g(x+b_0u)\\
=\;& b_0\left(m^{(1)}(x)\,u^T+b_0u\int_0^1(1-t)\,m^{(2)}(x+tb_0u)\,dtu^T\right)\left(g(x)+b_0\int_0^1 g^{(1)}(x+tb_0u)\,dtu^T\right).
\end{aligned}$$

Substituting the order in the bound above and (C.19) in (C.18), we obtain

$$V_n=\frac{O_{\mathbb{P}}(1)}{\left(nb_0^d\right)^{p_1}}\left[\left(n^{p_1}b_0^d\right)^{1/2}b_0^{p_1}+\left(nb_0^{d+2}\right)^{p_1}\right]=O_{\mathbb{P}}\left(b_0^{2p_1}\right),$$

since under (H_9), we have $b_0^d/(nb_0^{2d})^p = O(b_0^{2p})$, for all p in $[0,6]$. This proves (B.4).

Let now turn to (B.5). The Hölder, the Marcinkiewicz-Zygmund and the Minkowski inequalities give

$$\begin{aligned}
&\sum_{i=1}^{n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\mathbb{E}_n\left|\Sigma_{in}^{p_1}(x)K_1^{p_2}\left(\frac{X_i-x}{b_1}\right)\right|dx \\
&= \left|\sum_{i=1}^{n}\int \frac{\mathbb{1}\left(x\in\mathcal{X}\right)}{\left|nb_0^d\widehat{g}_n\left(x\right)\right|^{p_1}}\mathbb{E}_n\left[\left|\sum_{1\le j\ne i\le n}\varepsilon_j K_0\left(\frac{X_j-x}{b_0}\right)\right|^{p_1}\right]\left|K_1^{p_2}\left(\frac{X_i-x}{b_1}\right)\right|dx\right| \\
&\le \sum_{i=1}^{n}\int \frac{\mathbb{1}\left(x\in\mathcal{X}\right)}{\left|nb_0^d\widehat{g}_n\left(x\right)\right|^{p_1}}\mathbb{E}_n^{1/2}\left[\left|\sum_{1\le j\ne i\le n}\varepsilon_j K_0\left(\frac{X_j-x}{b_0}\right)\right|^{2p_1}\right]\left|K_1^{p_2}\left(\frac{X_i-x}{b_1}\right)\right|dx \\
&\le C\sum_{i=1}^{n}\int \frac{\mathbb{1}\left(x\in\mathcal{X}\right)}{\left|nb_0^d\widehat{g}_n\left(x\right)\right|^{p_1}}\left\{\mathbb{E}_n^{1/p_1}\left[\left|\sum_{1\le j\ne i\le n}\varepsilon_j^2 K_0^2\left(\frac{X_j-x}{b_0}\right)\right|^{p_1}\right]\right\}^{p_1/2}\left|K_1^{p_2}\left(\frac{X_i-x}{b_1}\right)\right|dx \\
&\le C\sum_{i=1}^{n}\int \frac{\mathbb{1}\left(x\in\mathcal{X}\right)}{\left|nb_0^d\widehat{g}_n\left(x\right)\right|^{p_1}}\left\{\sum_{1\le j\ne i\le n}\mathbb{E}_n^{1/p_1}\left[\left|\varepsilon_j^2 K_0^2\left(\frac{X_j-x}{b_0}\right)\right|^{p_1}\right]\right\}^{p_1/2}\left|K_1^{p_2}\left(\frac{X_i-x}{b_1}\right)\right|dx \\
&\le C\sum_{i=1}^{n}\int \frac{\mathbb{1}\left(x\in\mathcal{X}\right)}{(nb_0^d)^{p_1/2}\left|\widehat{g}_n\left(x\right)\right|^{p_1}}\left\{\frac{1}{nb_0^d}\sum_{1\le j\ne i\le n}\left|K_0\left(\frac{X_j-x}{b_0}\right)\right|\right\}^{p_1/2}\left|K_1^{p_2}\left(\frac{X_i-x}{b_1}\right)\right|dx.
\end{aligned}$$

It then follows from Lemma 4.7 that

$$\begin{aligned}
\sum_{i=1}^{n}\int \mathbb{1}\left(x\in\mathcal{X}\right)\mathbb{E}_n\left|\Sigma_{in}^{p_1}(x)K_1^{p_2}\left(\frac{X_i-x}{b_1}\right)\right|dx &= O_{\mathbb{P}}\left(\frac{nb_1^d}{\left(nb_0^d\right)^{p_1/2}}\right)\sup_{x\in\mathcal{X}}\frac{1}{nb_1^d}\sum_{i=1}^{n}\left|K_1^{p_2}\left(\frac{X_i-x}{b_1}\right)\right| \\
&= O_{\mathbb{P}}\left(\frac{nb_1^d}{\left(nb_0^d\right)^{p_1/2}}\right).
\end{aligned}$$

This proves (B.5) and completes the proof of the Lemma. □

Chapitre 5

Simulation study

Abstract : In this chapter we present our numerical results. We analyze and compare the performances of the Kernel density estimator $\widehat{f}_{1n}$, based on the estimated residuals, and the ones of the integral Kernel estimator $\widehat{f}_{2n}$. This comparison is made in the univariate case with a quadratic model, as described in the next section. The chapter is organized as follows. Section 5.1 is devoted to the description of our simulation framework. Section 5.2 investigates the global study for the estimators $\widehat{f}_{1n}$ and $\widehat{f}_{2n}$. We compare in that section the performances of these estimators in the sense of the Average Integrated Squared Error (AISE). Section 5.3 deals with the pointwise study of our two Kernel estimators, and compare their Average Squared Error (ASE), while section 5.4 investigates their asymptotic normality.

5.1 Description of our simulation framework

Let us consider the following quadratic model

$$Y = 3X^2 + 2X + 1 + \varepsilon, \tag{5.1.1}$$

where $\varepsilon \sim \mathrm{N}(0,1)$ and $X \sim \mathrm{U}[-1,1]$. For our numerical study, we generate $T = 100$ independent samples $(X_{k1}, \varepsilon_{k1}), (X_{k2}, \varepsilon_{k2}), \ldots, (X_{kn}, \varepsilon_{kn})$, $k = 1, \ldots, T$, of size $n = 200$, from the model (5.1.1). Define, for any integer $i \in [1, 200]$ and any integer $k \in [1, 100]$,

$$Y_{ki} = 3X_{ki}^2 + 2X_{ki} + 1 + \varepsilon_{ki}.$$

We denote by $\widehat{f^*}_{1k}(\epsilon)$ and $\widehat{f^*}_{2k}(\epsilon)$ the simulated versions of the estimators $\widehat{f}_{jn}(\epsilon)$ $(j = 1, 2)$ based on k^{th} sample $(X_{k1}, \varepsilon_{k1}), (X_{k2}, \varepsilon_{k2}), \ldots, (X_{kn}, \varepsilon_{kn})$. Hence the estimators $\widehat{f}_{jn}(\epsilon)$ are approximated by

$$\overline{\widehat{f}}_{jn}(\epsilon) = \frac{1}{T} \sum_{k=1}^{T} \widehat{f^*}_{jk}(\epsilon), \quad j = 1, 2. \tag{5.1.2}$$

For the estimator $\widehat{f}_{1n}(\epsilon)$, we do not make a truncation and consider $\mathcal{X}_0 = [-1, 1]$ in the estimator $\widehat{f}_{1n}(\epsilon)$. We also denote by $\widetilde{f}_{1n}(\epsilon)$ the Kernel estimator of $f(\epsilon)$ based on the true residuals, and by

$\widetilde{f}_{2n}(\epsilon)$ the integral Kernel estimator of $f(\epsilon)$ based on the true regression function. That is,

$$\begin{aligned}
\widetilde{f}_{1n}(\epsilon) &= \frac{1}{nb_1}\sum_{i=1}^{n} K_1\left(\frac{\varepsilon_i-\epsilon}{b_1}\right), \\
\widetilde{f}_{2n}(\epsilon) &= \int_{-1}^{1} \widehat{\varphi}_n\left(x, \epsilon+m(x)\right) dx,
\end{aligned}$$

where $m(x) = 3x^2 + 2x + 1$, and $\widehat{\varphi_n}$ is defined as in Chapter 4. Hence we can approximate these estimators by

$$\overline{\widetilde{f}}_{jn}(\epsilon) = \frac{1}{T}\sum_{k=1}^{T} \widetilde{f^*}_{jk}(\epsilon), \quad j = 1, 2, \tag{5.1.3}$$

where $\widetilde{f^*}_{jk}(\epsilon)$ is the Kernel version of $\widetilde{f}_{jn}(\epsilon)$ based on the k^{th} generated sample.

For the choice of the Kernels functions K_ℓ, $\ell = 0, 1, 2$, we consider the Epanechnikov Kernel function

$$K(x) = K_0(x) = K_1(x) = \frac{3}{4}\left(1-x^2\right) \mathbb{1}\left(|x| \le 1\right),$$

and the the biquadratic or biweight Kernel function

$$K_2(x) = \frac{15}{16}\left(1-x^2\right)^2 \mathbb{1}\left(|x| \le 1\right).$$

Recall that the numerical value of $\widetilde{f}_{2n}(\epsilon)$ is approximated by the Riemann sum

$$S_n(\epsilon) = \sum_{j=1}^{p} \widehat{\varphi}_n\left(x_j, \epsilon + m(x_j)\right)\left(x_j - x_{j-1}\right),$$

where $\{x_0, x_1, \ldots, x_p\}$ is a set of points such that $-1 = x_0 < x_1 < \ldots < x_p = 1$. In our setup, the sequence (x_j) is chosen such that $p = 100$ and

$$x_j = -1 + \frac{2j}{p}, \quad j = 1, \ldots, p.$$

5.2 Global study

In the nonparametric density estimation, it is known that a proper choice of the bandwidths is crucial for the precision of the estimator. In our simulations setup, we first find the simulated optimal bandwidths for the estimators $\widehat{f}_{jn}$, $j = 1, 2$. To that aim, we need to apply the Mean Itegrated Square Error (MISE) criterion which consists to minimize the quantities

$$\text{MISE}(\widehat{f}_{jn}) = \mathbb{E}\left[\int_{-A}^{A}\left(\widehat{f}_{jn}(t) - f(t)\right)^2 dt\right],$$

where $[-A, A]$ is a set that contains all the simulated residuals. In this subsection, we suppose that $[-A, A] = [-5, 5]$. For the sake of simplicity, we assume that $h = b_1$ for the estimator $\widehat{f}_{2n}(\epsilon)$. Using the T generated samples, we can approximate the MISE of the estimators $\widehat{f}_{jn}$ by the simulated Average Integrated Square Error (AISE) defined as follows :

$$\text{AISE}(\widehat{f}_{jn}) = \text{AISE}(\widehat{f}_{jn})(b_1, b_0) = \frac{1}{T}\sum_{k=1}^{T}\int\left(\widehat{f^*}_{jk}(t) - f(t)\right)^2 dt.$$

Now for each j, we denote by $(\widehat{b}_{1j}, \widehat{b}_{0j})$ the optimal bandwidths that minimize the above AISE. These bandwidths are simulated from $T = 100$ other independent samples of size $n = 200$ generated from the model (5.1.1), and different of the samples $(X_{k1}, \varepsilon_{k1}), (X_{k2}, \varepsilon_{k2}), \ldots, (X_{kn}, \varepsilon_{kn})$.

In Figures 5.1 and 5.2, we plot the AISE of the estimators $\widehat{f}_{1n}$ and $\widehat{f}_{2n}$ when b_0 and b_1 vary on $[0.1, 1.1]$ in the set $\{h_j = 0.1 + (0.01) \times j, 1 \leq j \leq 100\}$. The first plot shows that the optimal bandwidths $(\widehat{b}_{11}, \widehat{b}_{01})$ for the Kernel estimator $\widehat{f}_{1n}$ would be achieved when the couple $(\widehat{b}_{11}, \widehat{b}_{01})$ is very close to $(1, 0.2)$, while the second plot reveals that $(\widehat{b}_{12}, \widehat{b}_{02})$ should be achieved at the neighborhood of $(0.2, 0.2)$.

These graphical results about the bandwidths $(\widehat{b}_{1j}, \widehat{b}_{0j})$ are confirmed by the numerical results of Table 5.1, in which we give the optimal bandwidths for estimators $\widehat{f}_{jn}$ and $\widetilde{f}_{jn}$, and their corresponding AISE. For the $(\widehat{b}_{1j}, \widehat{b}_{0j})$, we observe that $\widehat{b}_{01}$ is approximately as small as $\widehat{b}_{02}$, while $\widehat{b}_{11}$ and $\widehat{b}_{12}$ are clearly different, same as $\widetilde{b}_1$ and $\widetilde{b}_2$. The results of Table 5.1 also reveal that $\text{AISE}(\widehat{f}_{1n})(\widehat{b}_{11}, \widehat{b}_{01}) < \text{AISE}(\widehat{f}_{2n})(\widehat{b}_{12}, \widehat{b}_{02})$, $\text{AISE}(\widehat{f}_{2n})$ being approximately twice as big as $\text{AISE}(\widehat{f}_{1n})$. This would suggest that for a judicious choice of the bandwidths (b_0, b_1), the AISE of the estimator $\widehat{f}_{1n}$ is smaller than the one of $\widehat{f}_{2n}$. Consequently $\widehat{f}_{1n}$ should be preferred to $\widehat{f}_{2n}$ for the estimation of p.d.f of the residuals. Moreover, Table 5.1 shows that $\text{AISE}(\widehat{f}_{1n})(\widehat{b}_{11}, \widehat{b}_{01}) \approx \text{AISE}(\widetilde{f}_{1n})(\widetilde{b}_1)$ and that $\text{AISE}(\widehat{f}_{2n})(\widehat{b}_{12}, \widehat{b}_{02}) < \text{AISE}(\widetilde{f}_{2n})(\widetilde{b}_2)$.

FIGURE 5.1 – The AISE of the Kernel estimator $\widehat{f}_{1n}$ based on the estimated residuals.

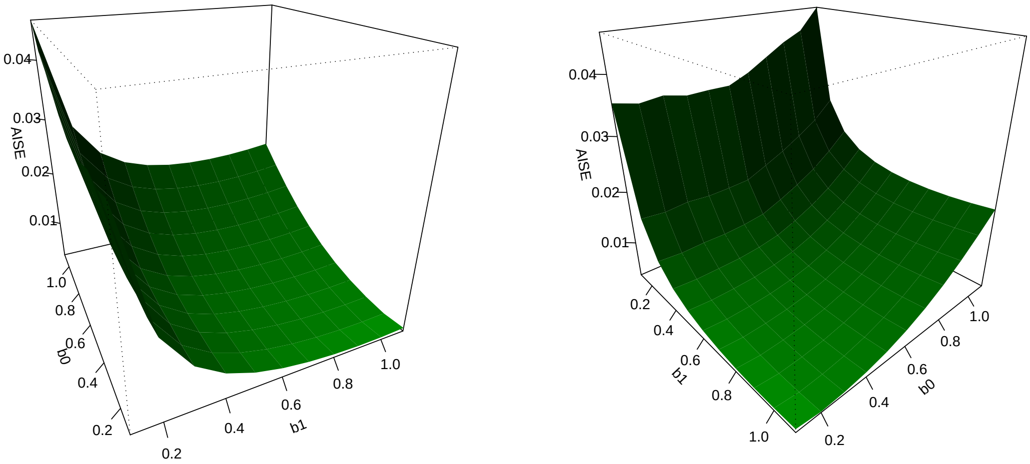


FIGURE 5.2 – The AISE of the integral Kernel estimator $\widehat{f}_{2n}$.

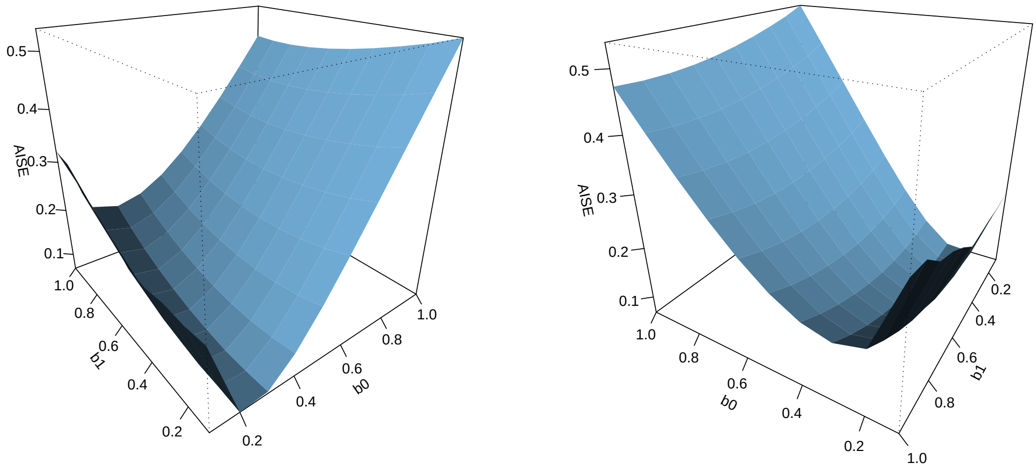


TABLE 5.1 – The optimal bandwidths $(\widehat{b}_{1j}, \widehat{b}_{0j})$, $\widetilde{b}_j$ and their corresponding AISE when b_0 and b_1 vary on $[0.1, 1.1]$.

$\widehat{f}_{1n}$			$\widetilde{f}_{1n}$		$\widehat{f}_{2n}$			$\widetilde{f}_{2n}$	
$\widehat{b}_{11}$	$\widehat{b}_{01}$	AISE$(\widehat{b}_{11}, \widehat{b}_{01})$	$\widetilde{b}_1$	AISE$(\widetilde{b}_1)$	$\widehat{b}_{12}$	$\widehat{b}_{02}$	AISE$(\widehat{b}_{12}, \widehat{b}_{02})$	$\widetilde{b}_2$	AISE$(\widetilde{b}_2)$
0.95	0.19	0.003141035	1.01	**0.003083492**	0.24	0.17	**0.006217096**	0.22	0.006406112

Table 5.1 shows that the optimal first-step bandwidths for the estimators $\widehat{f}_{1n}$ and $\widehat{f}_{2n}$ would be very small, as recommended in Wang, Brown, Cai and Levine (2008).

We now define for $\alpha = 0.05$ and $\alpha = 0.95$, the αth confidence band $\widehat{f}_{jn}(\cdot, \alpha)$ of the estimator $\widehat{f}_{jn}(\cdot)$ as follows. For each j and any $\epsilon \in [-5, 5]$, we consider the T ordered values $\widehat{f^*}_{j,(k)}(\epsilon)$ of the $\widehat{f^*}_{jk}(\epsilon)$'s such that $\widehat{f^*}_{j,(1)}(\epsilon) \leq \widehat{f^*}_{j,(2)}(\epsilon) \leq \ldots \leq \widehat{f^*}_{j,(T)}(\epsilon)$. Hence the function $\widehat{f}_{jn}(\alpha, \cdot)$ is defined as

$$\widehat{f}_{jn}(\epsilon, \alpha) = \widehat{f^*}_{j,(\alpha T)}(\epsilon), \quad \epsilon \in [-5, 5].$$

Using the optimal bandwidths $(\widehat{b}_{1j}, \widehat{b}_{0j})$ described above, we represent in Figures 5.3 and 5.4 the Average Kernel estimators $\overline{\widehat{f}}_{jn}$, the p.d.f of $N(0, 1)$, the 0.95th and the 0.05th confidence bands of the estimators $\widehat{f}_{jn}$, $j = 1, 2$. These plots can be useful for having a general idea about the confidence interval of the density f. For example, we see that for ϵ varying in the neighborhood of 0, we have $\widehat{f}_{jn}(\epsilon, 0.05) < f(\epsilon) < \widehat{f}_{jn}(\epsilon, 0.95)$.

In each of the Figures 5.3 and 5.4, the bias of the estimated density is quite important around the inflexion point $\epsilon = 0$, but the true density function remains in the good confidence interval. We also notice that the graphics plotted in Figure 5.4 are less smooth than the ones represented in Figure 5.3. This may explain the fact that AISE$(\widehat{f}_{1n})(\widehat{b}_{11}, \widehat{b}_{01}) <$ AISE$(\widehat{f}_{2n})(\widehat{b}_{12}, \widehat{b}_{02})$.

FIGURE 5.3 – From top to bottom, the 0.95th confidence band of $\widehat{f}_{1n}$, the p.d.f of N(0,1), the Average Kernel estimator $\overline{\widehat{f}}_{1n}$ and the 0.05th confidence band of $\widehat{f}_{1n}$ when $b_1 = \widehat{b}_{11} = 0.95$ and $b_0 = \widehat{b}_{01} = 0.19$.

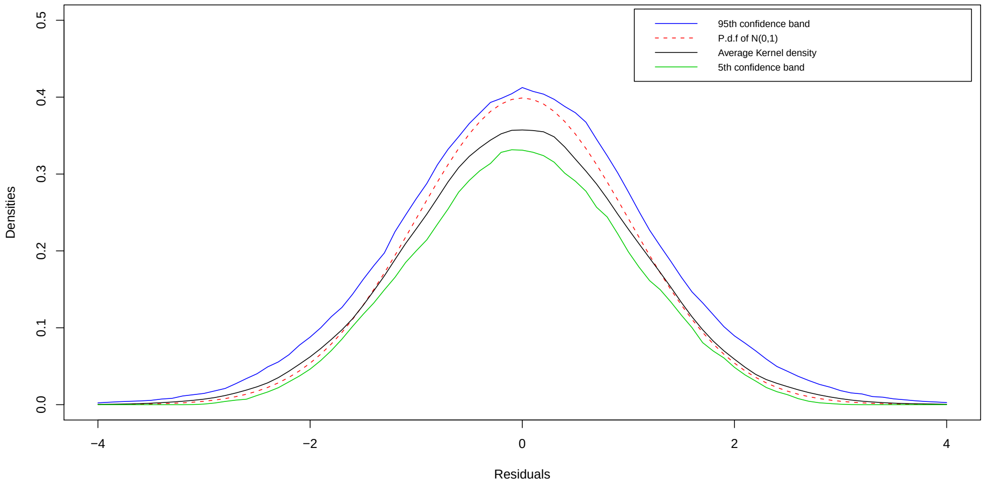
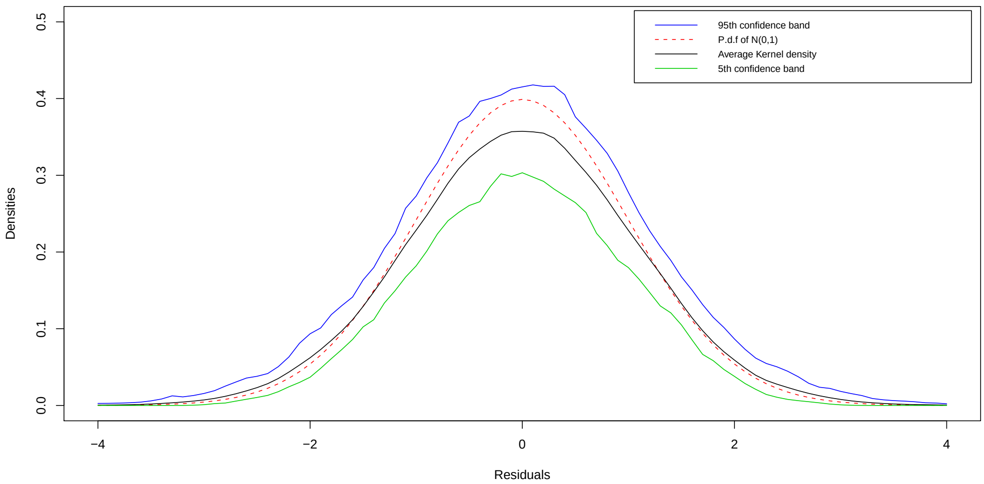


FIGURE 5.4 – From top to bottom, the 0.95th confidence band of $\widehat{f}_{2n}$, the p.d.f of N(0,1), the Average Kernel estimator $\overline{\widehat{f}}_{2n}$ and the 0.05th confidence band of $\widehat{f}_{2n}$ when $b_1 = \widehat{b}_{12} = 0.24$ and $b_0 = \widehat{b}_{02} = 0.12$.



5.3 Pointwise study

In this section, we are interested in the pointwise study of the estimators $\widehat{f}_{jn}(\epsilon)$ and $\widetilde{f}_{jn}(\epsilon)$. First, we compare the Average Square Errors (ASE) of these estimators at the points $\epsilon = -1, 0, 1$. In a second time, a comparison of the bias and variances of these estimators is established, and next their asymptotic normality is investigated.

5.3.1 Comparison of the ASE

Let $(\widehat{f}_{jn}(\epsilon), \widehat{f^*}_{jk}(\epsilon))$ and $(\widetilde{f}_{jn}(\epsilon), \widetilde{f^*}_{jk}(\epsilon))$ be as in the previous subsection. We compare the pointwise ASE of the estimators $\widehat{f}_{jn}(\epsilon)$ to the ones of the estimators $\widetilde{f}_{jn}(\epsilon)$. These ASE are defined as

$$\begin{aligned} \mathrm{ASE}(\widehat{f}_{jn})(\epsilon) &= \frac{1}{T}\sum_{k=1}^{T}\left(\widehat{f^*}_{jk}(\epsilon) - f(\epsilon)\right)^2, \\ \mathrm{ASE}(\widetilde{f}_{jn})(\epsilon) &= \frac{1}{T}\sum_{k=1}^{T}\left(\widetilde{f^*}_{jk}(\epsilon) - f(\epsilon)\right)^2. \end{aligned}$$

The comparison of the ASE is done at the points $\epsilon = -1, 0, 1$, using respectively the pointwise optimal bandwidths

$$(\widehat{b}_{1j}(\epsilon), \widehat{b}_{0j}(\epsilon)) = \arg\min_{(b_1, b_0)} \mathrm{ASE}(\widehat{f}_{jn})(\epsilon), \quad \widetilde{b}_j(\epsilon) = \arg\min_{(b_1, b_0)} \mathrm{ASE}(\widetilde{f}_{jn})(\epsilon).$$

As in the global study, these bandwidths are based upon $T = 100$ new independent samples of size $n = 200$ generated from the model (5.1.1), and different of the samples that are used for computing $\mathrm{ASE}(\widehat{f}_{jn})(\epsilon, b_1, b_0)$ and $\mathrm{ASE}(\widetilde{f}_{jn})(\epsilon, b_1)$. In this section, the minimizations of the ASE are performed for b_1 and b_0 varying on $[0.1, 3]$, in the set $\{h_j = 0.1 + (0.01) \times j, 1 \le j \le 290\}$. For $j = 1, 2$ and $\epsilon = -1, 0, 1$, the optimal values of the $\mathrm{ASE}(\widehat{f}_{jn})(\epsilon)$ and $\mathrm{ASE}(\widetilde{f}_{jn})(\epsilon)$ are gathered in Tables 5.2 and 5.3. These values show that for any $\epsilon = -1, 0, 1$,

$$\mathrm{ASE}(\widehat{f}_{1n})(\epsilon, \widehat{b}_{11}, \widehat{b}_{01}) < \mathrm{ASE}(\widehat{f}_{2n})(\epsilon, \widehat{b}_{12}, \widehat{b}_{02}). \tag{5.3.4}$$

This fact parallels the results of the Global study in which we saw that for an optimal choice of the bandwidth, the AISE of the estimator $\widehat{f}_{1n}$ is smaller than the one of the estimator $\widehat{f}_{2n}$. Consequently the pointwise estimator $\widehat{f}_{1n}(\epsilon)$ should also be preferred to the estimator $\widehat{f}_{2n}(\epsilon)$ for the nonparametric Kernel estimation of $f(\epsilon)$.

TABLE 5.2 – ASE of $\widehat{f}_{1n}(\epsilon)$ and $\widetilde{f}_{1n}(\epsilon)$ using the bandwidths $(\widehat{b}_{11}(\epsilon), \widehat{b}_{01}(\epsilon))$ and $\widetilde{b}_1(\epsilon)$.

$\epsilon = -1$		$\epsilon = 0$		$\epsilon = 1$	
$\mathrm{ASE}(\widehat{f}_{1n})$	$\mathrm{ASE}(\widetilde{f}_{1n})$	$\mathrm{ASE}(\widehat{f}_{1n})$	$\mathrm{ASE}(\widetilde{f}_{1n})$	$\mathrm{ASE}(\widehat{f}_{1n})$	$\mathrm{ASE}(\widetilde{f}_{1n})$
0.00020536762	0.00023502221	0.0015443395	**0.0011523854**	**0.00013607338**	0.00028682107

TABLE 5.3 – ASE of $\widehat{f}_{2n}(\epsilon)$ and $\widetilde{f}_{2n}(\epsilon)$ based on the bandwidths $(\widehat{b}_{12}(\epsilon), \widehat{b}_{02}(\epsilon))$ and $\widetilde{b}_2(\epsilon)$.

$\epsilon = -1$		$\epsilon = 0$		$\epsilon = 1$	
$\mathrm{ASE}(\widehat{f}_{2n})$	$\mathrm{ASE}(\widetilde{f}_{2n})$	$\mathrm{ASE}(\widehat{f}_{2n})$	$\mathrm{ASE}(\widetilde{f}_{2n})$	$\mathrm{ASE}(\widehat{f}_{2n})$	$\mathrm{ASE}(\widetilde{f}_{2n})$
0.00086381543	0.0009047470	0.0026912672	**0.0025553406**	**0.00092628778**	0.0014950721

From Tables 5.2 and 5.3, we also notice that for $j = 1, 2$,

$$\mathrm{ASE}(\widehat{f}_{jn})(0) \approx \mathrm{ASE}(\widetilde{f}_{jn})(0), \quad \mathrm{ASE}(\widehat{f}_{jn})(\epsilon) < \mathrm{ASE}(\widetilde{f}_{jn})(\epsilon), \quad \epsilon = -1, 1.$$

For the estimation of linear functionals of the error distribution in a semiparametric context, Müller, Schick and Wefelmeyer (2004) have shown that the estimators using the estimated residuals may have a smaller asymptotic variance compared to estimators that are based on the true errors. A reason that may explain this effect is that the estimators $\widetilde{f}_{jn}(\epsilon)$ do not use the fact that the residuals ε_i have mean zero, contrarily to the estimators $\widehat{f}_{jn}(\epsilon)$. Note however that the improvement of $\widehat{f}_{1n}(\epsilon)$ on $\widetilde{f}_{1n}(\epsilon)$ is much more clear-cut in the pointwise setup than in the global one.

Nevertheless, we observe that for the estimator $\widehat{f}_{1n}$ based on the estimated residuals, the values of the ASE are quite different at the points $\epsilon = -1$ and $\epsilon = 1$. We then attempt to explain this situation by analyzing the behavior of the error terms around these points. Define, for any integers $k \in [1, T]$ and $i \in [1, n]$,

$$\widehat{\delta}_{ki}(\epsilon) = (\widehat{\varepsilon}_{ki} - \varepsilon_{ki})\, \mathbb{1}\left(|\widehat{\varepsilon}_{ki} - \epsilon| \leq \widehat{b}_{11}(\epsilon)\right),$$

where $\widehat{\varepsilon}_{ki}$ is the Kernel empirical version of ε_{ki} based on the optimal first-step bandwidth $\widehat{b}_{01}(\epsilon)$ for the estimator $\widehat{f}_{1n}(\epsilon)$. We then define the empirical mean $\overline{\delta}(\epsilon)$ and the empirical variance $\sigma^2_{\overline{\delta}}(\epsilon)$ of the $\widehat{\delta}_{ki}(\epsilon)$'s as

$$\overline{\delta}(\epsilon) = \frac{1}{nT} \sum_{k=1}^{T} \sum_{i=1}^{n} \widehat{\delta}_{ki}(\epsilon), \quad \sigma^2_{\overline{\delta}}(\epsilon) = \frac{1}{nT} \sum_{k=1}^{T} \sum_{i=1}^{n} \left(\widehat{\delta}_{ki}(\epsilon) - \overline{\delta}(\epsilon)\right)^2.$$

TABLE 5.4 – Values of the empirical means $\overline{\delta}(\epsilon)$ and the empirical variances $\sigma^2_{\overline{\delta}}(\epsilon)$ for $\epsilon = -1, 1$.

$\epsilon = -1$		$\epsilon = 1$	
$\overline{\delta}(\epsilon)$	$\sigma^2_{\overline{\delta}}(\epsilon)$	$\overline{\delta}(\epsilon)$	$\sigma^2_{\overline{\delta}}(\epsilon)$
-0.3911658	0.3331359	0.03403744	0.04659867

In Table 5.4, we evaluate the quantities $\overline{\delta}(\epsilon)$ and $\sigma^2_{\overline{\delta}}(\epsilon)$, using the bandwidths $b_0 = \widehat{b}_{01}(\epsilon)$ and $b_1 = \widehat{b}_{11}(\epsilon)$, $\epsilon = -1, 1$. We observe that the variables $\widehat{\delta}_{ki}(-1)$ have a lower empirical bias and a higher empirical variance than the data $\widehat{\delta}_{ki}(1)$. Hence around the point $\epsilon = -1$, the error percentage for the estimation of the true residuals ε_{ki} by the nonparametric residuals $\widehat{\varepsilon}_{ki}$ is more important than around the point $\epsilon = 1$. This may explain the difference of the ASE at the points $\epsilon = -1, 1$ for the estimator $\widehat{f}_{1n}$, as seen in Table 5.2.

5.3.2 Comparison of the bias and variances

In this subsection, we suppose that the estimators $\widehat{f^*}_{jk}(\epsilon)$ and $\widetilde{f^*}_{jk}(\epsilon)$ $(j = 1, 2)$ defined in the previous subsection are respectively based upon the optimal bandwidths $(\widehat{b}_{1j}(\epsilon), \widehat{b}_{0j}(\epsilon))$ and

$\widetilde{b}_j(\epsilon)$. For each j, let $\widehat{B}_{jn}(\epsilon)$ and $\widetilde{B}_{jn}(\epsilon)$ be respectively the empirical bias of the estimated densities $\widehat{f}_{jn}(\epsilon)$ and $\widetilde{f}_{jn}(\epsilon)$. These estimated quantities are defined as

$$\widehat{B}_{jn}(\epsilon) = \frac{1}{T}\sum_{k=1}^{T}\left(\widehat{f^*}_{jk}(\epsilon) - f(\epsilon)\right), \quad \widetilde{B}_{jn}(\epsilon) = \frac{1}{T}\sum_{k=1}^{T}\left(\widetilde{f^*}_{jk}(\epsilon) - f(\epsilon)\right).$$

The simulated values of the bias $\widehat{B}_{jn}(\epsilon)$ and $\widetilde{B}_{jn}(\epsilon)$ at the points $\epsilon = -1, 0, 1$ are represented in Table 5.5 and 5.6.

TABLE 5.5 – Optimal values of the bias $\widehat{B}_{1n}(\epsilon)$ and $\widetilde{B}_{1n}(\epsilon)$.

$\epsilon = -1$		$\epsilon = 0$		$\epsilon = 1$	
$\widehat{B}_{1n}(\epsilon)$	$\widetilde{B}_{1n}(\epsilon)$	$\widehat{B}_{1n}(\epsilon)$	$\widetilde{B}_{1n}(\epsilon)$	$\widehat{B}_{1n}(\epsilon)$	$\widetilde{B}_{1n}(\epsilon)$
-0.005290204	-0.008126554	-0.02450615	**-0.01726208**	**-0.005446647**	-0.008392434

TABLE 5.6 – Optimal values of the bias $\widehat{B}_{2n}(\epsilon)$ and $\widetilde{B}_{2n}(\epsilon)$.

$\epsilon = -1$		$\epsilon = 0$		$\epsilon = 1$	
$\widehat{B}_{2n}(\epsilon)$	$\widetilde{B}_{2n}(\epsilon)$	$\widehat{B}_{2n}(\epsilon)$	$\widetilde{B}_{2n}(\epsilon)$	$\widehat{B}_{2n}(\epsilon)$	$\widetilde{B}_{2n}(\epsilon)$
-0.01615247	-0.01661985	**-0.02447482**	-0.02612883	-0.01803243	**-0.01262712**

Table 5.5 reveals that $|\widehat{B}_{1n}(\epsilon)| < |\widehat{B}_{2n}(\epsilon)|$ for $\epsilon = -1$ and $\epsilon = 1$, and that $|\widehat{B}_{1n}(0)| \approx |\widehat{B}_{2n}(0)|$. This indicates that at the points $\epsilon = -1$ and $\epsilon = 1$, the estimator $\widehat{f}_{1n}(\epsilon)$ would be less biased than the estimator $\widehat{f}_{2n}(\epsilon)$.

Moreover for $\epsilon = -1$ and $\epsilon = 1$, the estimator $\widehat{f}_{1n}(\epsilon)$ is much less biased than the estimator $\widetilde{f}_{1n}(\epsilon)$. Consequently, there is a positive influence of the bandwidth $\widehat{b}_{01}(\epsilon)$ on the bias of $\widehat{f}_{1n}(\epsilon)$. But this situation contrasts with the one observed at $\epsilon = 0$, for which $\widehat{f}_{1n}(\epsilon)$ is more biased than $\widetilde{f}_{1n}(\epsilon)$.

For $\widehat{f}_{2n}(\epsilon)$ and $\widetilde{f}_{2n}(\epsilon)$, we note that the bias of these estimators are very close at the points $\epsilon = -1, 0, 1$. This means that the estimation of the regression function has a negligible impact on the bias of the estimator $\widehat{f}_{2n}(\epsilon)$.

Now, let $\widehat{V}_{jn}(\epsilon)$ and $\widetilde{V}_{jn}(\epsilon)$ be the estimated variances of $\widehat{f}_{jn}(\epsilon)$ and $\widetilde{f}_{jn}(\epsilon)$ defined as

$$\widehat{V}_{jn}(\epsilon) = \frac{1}{T}\sum_{k=1}^{T}\left(\widehat{f^*}_{jk}(\epsilon) - \widehat{\mu}_{jn}(\epsilon)\right)^2, \quad \widetilde{V}_{jn}(\epsilon) = \frac{1}{T}\sum_{k=1}^{T}\left(\widetilde{f^*}_{jk}(\epsilon) - \widetilde{\mu}_{jn}(\epsilon)\right)^2,$$

where

$$\widehat{\mu}_{jn}(\epsilon) = \frac{1}{T}\sum_{k=1}^{T}\widehat{f^*}_{jk}(\epsilon), \quad \widetilde{\mu}_{jn}(\epsilon) = \frac{1}{T}\sum_{k=1}^{T}\widetilde{f^*}_{jk}(\epsilon).$$

The simulated values of these empirical parameters are gathered in Tables 5.7 and 5.8.

TABLE 5.7 – Optimal values of variances $\widehat{V}_{1n}(\epsilon)$ and $\widetilde{V}_{1n}(\epsilon)$.

$\epsilon=-1$		$\epsilon=0$		$\epsilon=1$	
$\widehat{V}_{1n}(\epsilon)$	$\widetilde{V}_{1n}(\epsilon)$	$\widehat{V}_{1n}(\epsilon)$	$\widetilde{V}_{1n}(\epsilon)$	$\widehat{V}_{1n}(\epsilon)$	$\widetilde{V}_{1n}(\epsilon)$
0.0001773813	**0.0001689813**	0.0009437874	**0.0008544052**	**0.0001064074**	0.000216388

TABLE 5.8 – Optimal values of the variances $\widehat{V}_{2n}(\epsilon)$ and $\widetilde{V}_{2n}(\epsilon)$.

$\epsilon=-1$		$\epsilon=0$		$\epsilon=1$	
$\widehat{V}_{2n}(\epsilon)$	$\widetilde{V}_{2n}(\epsilon)$	$\widehat{V}_{2n}(\epsilon)$	$\widetilde{V}_{2n}(\epsilon)$	$\widehat{V}_{2n}(\epsilon)$	$\widetilde{V}_{2n}(\epsilon)$
0.0006029132	0.0006285278	0.00209225	**0.001872625**	**0.0006011193**	0.001335628

From Table 5.7, we notice that $\widehat{V}_{1n}(\epsilon) < \widehat{V}_{2n}(\epsilon)$ for $\epsilon = -1, 0, 1$. Consequently the estimator $\widehat{f}_{1n}(\epsilon)$ should be preferred to $\widehat{f}_{2n}(\epsilon)$, since the latter estimator is less efficient than the first one.

Moreover, we observe that $\widehat{V}_{1n}(\epsilon)$ is much less than $\widetilde{V}_{1n}(\epsilon)$ at $\epsilon = 1$, and slightly equal to $\widetilde{V}_{1n}(\epsilon)$ when $\epsilon = -1$ and $\epsilon = 0$. This means that the estimation of the residuals may have a positive influence on the final estimator $\widehat{f}_{1n}(\epsilon)$.

For the variances $\widehat{V}_{2n}(\epsilon)$ and $\widetilde{V}_{2n}(\epsilon)$, it is seen that the first variance is much less than the latter one at $\epsilon = 1$, and very close to $\widetilde{V}_{2n}(\epsilon)$ when $\epsilon = -1$ and $\epsilon = 0$. Hence the estimation of the regression function m may have a positive effect on the estimator $\widehat{f}_{2n}(\epsilon)$.

In conclusion, we note that at the points $\epsilon = -1, 0, 1$, the estimator $\widehat{f}_{1n}(\epsilon)$ dominates the estimator $\widehat{f}_{2n}(\epsilon)$ for the ASE, the bias and the variance. As in the Global study, this suggests that the first estimator should be preferred to the second one when we are interested in their Pointwise study.

5.3.3 Asymptotic normality

We examine here the asymptotic normality of the estimators $\widehat{f}_{jn}(\epsilon)$, for $j = 1, 2$ and $\epsilon = -1, 0, 1$. To that aim, we introduce the standardized variables

$$\widehat{Z}_{jn}(\epsilon) = \frac{\sqrt{n\widehat{b}_{1j}(\epsilon)}\left(\widehat{f}_{jn}(\epsilon) - f(\epsilon)\right)}{\sqrt{f(\epsilon)\int K_1^2(v)dv}}, \quad \widehat{Z^*}_{jk}(\epsilon) = \frac{\sqrt{n\widehat{b}_{1j}(\epsilon)}\left(\widehat{f^*}_{jk}(\epsilon) - f(\epsilon)\right)}{\sqrt{f(\epsilon)\int K_1^2(v)dv}}, \quad k = 1, \ldots, T,$$

where the $\widehat{f^*}_{jk}(\epsilon)$'s are defined as in previous subsection, for the evaluation of the bias and variances. The empirical mean $\overline{\overline{\mu}}_{jn}(\epsilon)$ and the empirical variance $\overline{\overline{\sigma}}^2_{jn}(\epsilon)$ of the $\widehat{Z}_{jn}(\epsilon)$'s are such that

$$\overline{\overline{\mu}}_{jn}(\epsilon) = \frac{1}{T}\sum_{k=1}^{T}\widehat{Z^*}_{jk}(\epsilon), \quad \overline{\overline{\sigma}}^2_{jn}(\epsilon) = \frac{1}{T}\sum_{k=1}^{T}\left(\widehat{Z^*}_{jk}(\epsilon) - \overline{\overline{\mu}}_{jn}(\epsilon)\right)^2.$$

Are the data $\widehat{Z}_{jn}(\epsilon)$ normal distributed ?

For each j and ϵ, we wish to test the hypothesis

$$\mathrm{H_{0j}}(\epsilon) : \widehat{Z}_{jn}(\epsilon) \sim N\left(\mu_j(\epsilon), \sigma_j^2(\epsilon)\right) \text{ versus } \mathrm{H_{1j}}(\epsilon) : \widehat{Z}_{jn}(\epsilon) \nsim N\left(\mu_j(\epsilon), \sigma_j^2(\epsilon)\right),$$

where the parameters $\mu_j(\epsilon)$ and $\sigma_j^2(\epsilon)$ are unknown and have to be estimated. The normality of the data $\widehat{Z}_{jn}(\epsilon)$ can be tested by an analytical method such as the Lilliefors method for the Kolmogorov-Smirnov test. Let us perform this Lilliefors test that the data $\widehat{Z}_{jn}(\epsilon)$ come from the normal distribution. For this, we denote by $\widehat{\mathrm{KS}}_j(\epsilon)$ and $\widehat{p}_j(\epsilon)$ respectively as the Kolmogorov-Smirnov statistic and the p-value of the above test. With the Lilliefors's method, the evaluation of the p-values $\widehat{p}_j(\epsilon)$ and the statistics $\widehat{\mathrm{KS}}_j(\epsilon)$ accounts for the estimations of $\mu_j(\epsilon)$ and $\sigma_j^2(\epsilon)$. For the characteristics and the properties of the KS or Lilliefors's test, see Massey (1951), Shorack and Wellner (1986), Dallal and Wilkinson (1986), Lehmann and Romano (1998), and Thode (2002). In Table 5.9 we have gathered the numerical values of the $\widehat{\mathrm{KS}}_j(\epsilon)$'s and the $\widehat{p}_j(\epsilon)$'s.

TABLE 5.9 – Values of the statistics $\widehat{\mathrm{KS}}_j(\epsilon)$ and the p-values $\widehat{p}_j(\epsilon)$ of the $\widehat{Z}_j(\epsilon)$'s.

$\epsilon = -1$				$\epsilon = 0$				$\epsilon = 1$			
$\widehat{\mathrm{KS}}_1(\epsilon)$	$\widehat{p}_1(\epsilon)$	$\widehat{\mathrm{KS}}_2(\epsilon)$	$\widehat{p}_2(\epsilon)$	$\widehat{\mathrm{KS}}_1(\epsilon)$	$\widehat{p}_1(\epsilon)$	$\widehat{\mathrm{KS}}_2(\epsilon)$	$\widehat{p}_2(\epsilon)$	$\widehat{\mathrm{KS}}_1(\epsilon)$	$\widehat{p}_1(\epsilon)$	$\widehat{\mathrm{KS}}_2(\epsilon)$	$\widehat{p}_2(\epsilon)$
0.0506	0.9598	0.0427	0.9933	0.0713	0.6891	0.0518	0.9516	0.0746	0.6347	0.0882	0.4176

The results of Table 5.9 show that the hypothesis on the normality of the data is accepted, since $\widehat{p}_j(\epsilon) > 0.05 = \alpha$ (a default value of the level of significance). **Hence according to the Lilliefors method, we can accept the fact that the data $\widehat{Z}_{jn}(\epsilon)$ come from a normal distribution.**

Beside the Lilliefors test, there exists a graphical method for investigating the normality of the data. This method is the normal Q-Q plots of the variables $\widehat{Z}_{jn}(\epsilon)$. The Q-Q plot provides a graphical way to determine the level of normality. If the data fall exactly along a reference line

(called the Henry's line), then the hypothesis on their normality can be receivable. If the empirical data deviate widely from this line, the data are non-normal. In Figures 5.5, 5.6 and 5.7, we represent the normal Q-Q plots of the data $\widehat{Z}_{1n}(\epsilon)$ and $\widehat{Z}_{2n}(\epsilon)$ for $\epsilon = -1, 0, 1$.

FIGURE 5.5 – From left to right : normal Q-Q plot of the data $Z_{1n}(-1)$ and $Z_{2n}(-1)$.

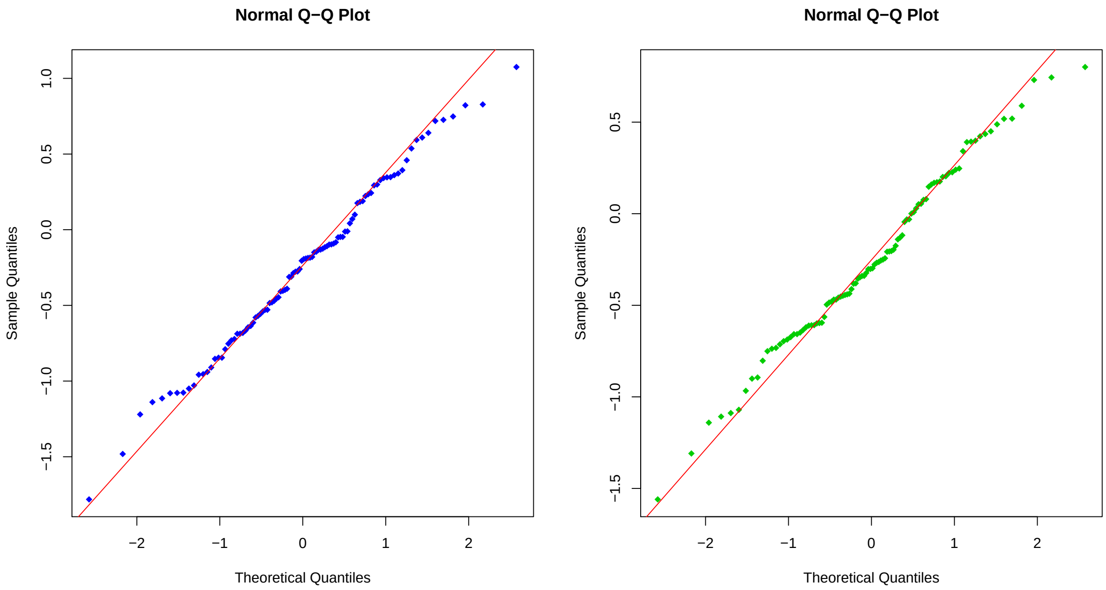


FIGURE 5.6 – Normal Q-Q plot of the data $Z_{1n}(0)$ and $Z_{2n}(0)$.

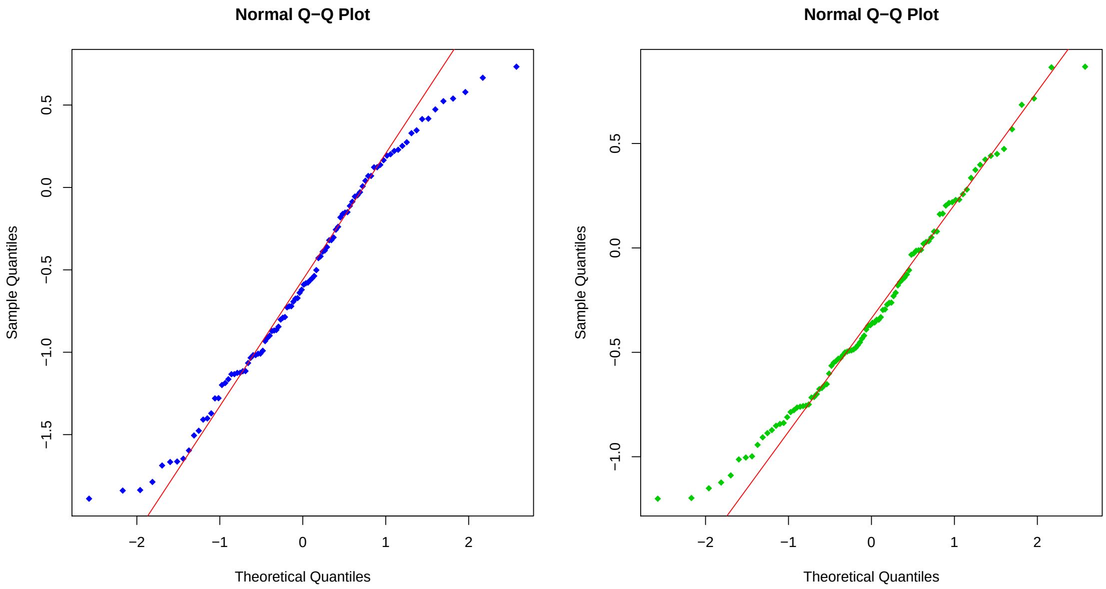
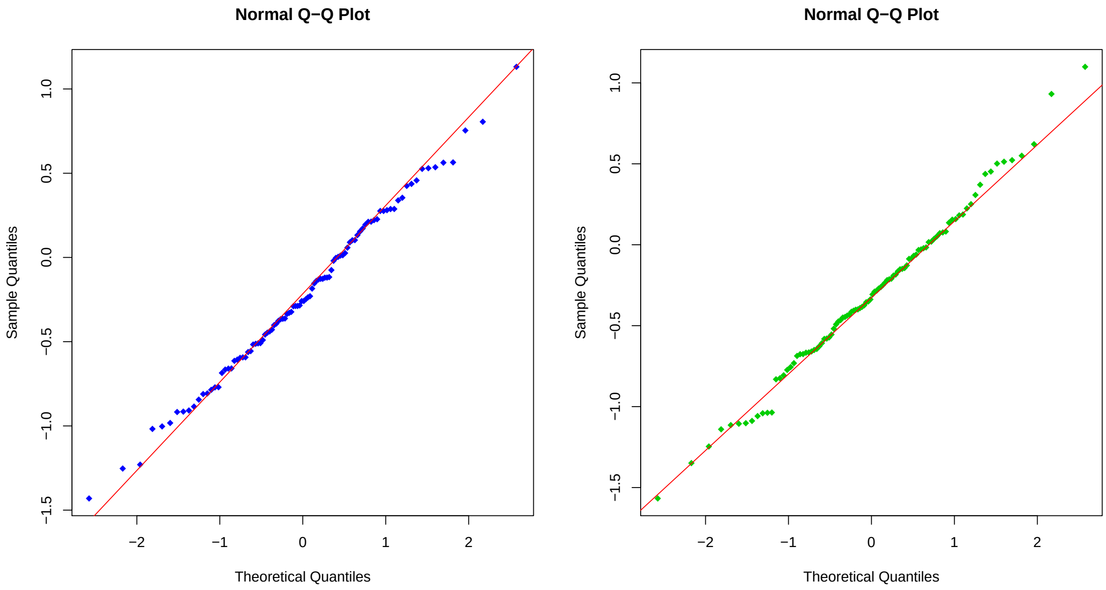


FIGURE 5.7 – Normal Q-Q plot of the data $Z_{1n}(1)$ and $Z_{2n}(1)$.



From the above figures, the hypothesis $\mathrm{H_{0j}}(\epsilon)$ can be receivable. However, we can have some doubts about the symmetry of the $\widehat{Z}_{jn}(\epsilon)$'s, since we notice that they deviate slightly from the Henry's line at the tails of the distribution. This augurs that the distribution of the variables $\widehat{Z}_{jn}(\epsilon)$ should not be symmetric.

Do the data $\widehat{Z}_{jn}(\epsilon)$ come from the standard normal $N(0,1)$?

Since the data $\widehat{Z}_{jn}(\epsilon)$ are standardized variables, we can wonder if they come from the normal distribution $N(0,1)$. To give some elements of answer to this question, we first compute the empirical bias and variances of the $\widehat{Z}_{jn}(\epsilon)$'s. These quantities are grouped in Table 5.10.

TABLE 5.10 – Values of the empirical means and variances of the $\widehat{Z}_{jn}(\epsilon)$'s.

$\epsilon=-1$				$\epsilon=0$				$\epsilon=1$			
$\overline{\overline{\mu}}_{1n}(\epsilon)$	$\overline{\overline{\sigma}}^2_{1n}(\epsilon)$	$\overline{\overline{\mu}}_{2n}(\epsilon)$	$\overline{\overline{\sigma}}^2_{2n}(\epsilon)$	$\overline{\overline{\mu}}_{1n}(\epsilon)$	$\overline{\overline{\sigma}}^2_{1n}(\epsilon)$	$\overline{\overline{\mu}}_{2n}(\epsilon)$	$\overline{\overline{\sigma}}^2_{2n}(\epsilon)$	$\overline{\overline{\mu}}_{1n}(\epsilon)$	$\overline{\overline{\sigma}}^2_{1n}(\epsilon)$	$\overline{\overline{\mu}}_{2n}(\epsilon)$	$\overline{\overline{\sigma}}^2_{2n}(\epsilon)$
-0.1459	0.4900	-0.3105	0.4098	-0.5798	0.6389	-0.3124	0.4380	-0.2491	0.4465	-0.3097	0.4323

This table shows that the empirical paremeters $\overline{\overline{\mu}}_{jn}(\epsilon)$ and $\overline{\overline{\sigma}}^2_{1n}(\epsilon)$ are clearly different to 0 and 1, which correspond respectively to the theoretical mean and variance of the normal $N(0,1)$.

We now evaluate the empirical quantiles of the variables $\widehat{Z}_{jn}(\epsilon)$. For each j, we consider the ordered values $\widehat{Z^*}_{j,(k)}(\epsilon)$ of the $\widehat{Z^*}_{jk}(\epsilon)$'s such that $\widehat{Z^*}_{j,(1)}(\epsilon) \leq \widehat{Z^*}_{j,(2)}(\epsilon) \leq \ldots \leq \widehat{Z^*}_{j,(T)}(\epsilon)$. Hence for any $\alpha \in [0,1]$, the α^{th} empirical quantiles of the $\widehat{Z}_{jn}(\epsilon)$'s are defined as $\widehat{Z}_{jn}(\epsilon,\alpha) = \widehat{Z^*}_{j,(\alpha T)}(\epsilon)$. In Table 5.11, we give the simulated values of these quantiles when $\alpha = 0.05$ and $\alpha = 0.95$.

TABLE 5.11 – Values of the quantiles $a_j(\epsilon) = \widehat{Z^*}_{j,(0.05\times T)}(\epsilon)$ and $c_j(\epsilon) = \widehat{Z^*}_{j,(0.95\times T)}(\epsilon)$.

$\epsilon=-1$				$\epsilon=0$				$\epsilon=1$			
$a_1(\epsilon)$	$c_1(\epsilon)$	$a_2(\epsilon)$	$c_2(\epsilon)$	$a_1(\epsilon)$	$c_1(\epsilon)$	$a_2(\epsilon)$	$c_2(\epsilon)$	$a_1(\epsilon)$	$c_1(\epsilon)$	$a_2(\epsilon)$	$c_2(\epsilon)$
-0.9719	0.6654	-0.9790	0.4006	-1.6874	0.4734	-1.0893	0.4738	-1.1377	0.5389	-1.1109	0.5071

From Table 5.11, we note first that the quantities $\widehat{Z}_{jn}(\epsilon, 0.05)$ and $\widehat{Z}_{jn}(\epsilon, 0.95)$ are globally clearly different to -1.64 and 1.64, the corresponding theoretical quantiles of the normal distribution $N(0,1)$. Next, the values of these quantiles show that the variables $\widehat{Z}_{jn}(\epsilon)$ are globally asymmetric. This suggests that the $\widehat{Z}_{jn}(\epsilon)$'s are not distributed according to the normal variable $N(0,1)$.

We now estimate the confidence intervals for the theoretical quantiles of the $\widehat{Z}_{jn}(\epsilon)$'s. This estimation is done under $\mathrm{H_{0j}}(\epsilon)$, and based on the following result. For $\alpha \in]0,1[$ and $T \to \infty$,

$$\frac{\widehat{Z^*}_{j,(\alpha T)(\epsilon)} - Q_{j,\alpha}(\epsilon)}{\sqrt{\alpha(\alpha-1)/\left(Tf^2\left(Q_{j,\alpha}(\epsilon)\right)\right)}} \xrightarrow{d} N(0,1),$$

where $Q_{j,\alpha}(\epsilon)$ is the theoretical α^{th} quantile of the variable $\widehat{Z}_{jn}(\epsilon)$, and $f(\cdot)$ the p.d.f of the normal variable $N(0,1)$. This result can be found, for example, in Tassi (1985). A consequence of a such result is that an asymptotic confidence interval for the $Q_{j,\alpha}(\epsilon)$'s, with a level of confidence $1-\alpha$, is given by

$$\widehat{I}_{j,\alpha}(\epsilon) = \left[\widehat{Z^*}_{j,(\alpha T)}(\epsilon) - \frac{q_{\alpha/2}\sqrt{\alpha(1-\alpha)}}{\sqrt{T}f\left(\widehat{Z^*}_{j,(\alpha T)}(\epsilon)\right)}, \widehat{Z^*}_{j,(\alpha T)}(\epsilon) + \frac{q_{\alpha/2}\sqrt{\alpha(1-\alpha)}}{\sqrt{T}f\left(\widehat{Z^*}_{j,(\alpha T)}(\epsilon)\right)}\right],$$

where $q_{\alpha/2}$ denotes the $(1-\alpha/2)$ quantile of the standard normal distribution. In Tables 5.12 and 5.13, we give the estimations of the $\widehat{I}_{j,\alpha}(\epsilon)$ when $\alpha = 0.05$ and $\alpha = 0.95$, with $T = 100$ and $n = 200$. As seen above in Table 5.11, the results of Tables 5.12 and 5.13 also reveal that the quantiles $Q_{j,0.05}(\epsilon)$ and $Q_{j,0.95}(\epsilon)$ should be respectively quite different to -1.64 and 1.64.

TABLE 5.12 – Confidence intervals of the theoretical quantiles $Q_{j,\alpha}(\epsilon)$ when $\alpha = 0.05$.

$\epsilon=-1$		$\epsilon=0$		$\epsilon=1$	
$\widehat{I}_{1,\alpha}(\epsilon)$	$\widehat{I}_{2,\alpha}(\epsilon)$	$\widehat{I}_{1,\alpha}(\epsilon)$	$\widehat{I}_{2,\alpha}(\epsilon)$	$\widehat{I}_{1,\alpha}(\epsilon)$	$\widehat{I}_{2,\alpha}(\epsilon)$
$[-1.143, -0.800]$	$[-1.152, -0.806]$	$[-2.132, -1.243]$	$[-1.283, -0.896]$	$[-1.342, -0.933]$	$[-1.309, -0.913]$

TABLE 5.13 – Confidence intervals of the theoretical quantiles $Q_{j,\alpha}(\epsilon)$ when $\alpha = 0.95$.

$\epsilon=-1$		$\epsilon=0$		$\epsilon=1$	
$\widehat{I}_{1,\alpha}(\epsilon)$	$\widehat{I}_{2,\alpha}(\epsilon)$	$\widehat{I}_{1,\alpha}(\epsilon)$	$\widehat{I}_{2,\alpha}(\epsilon)$	$\widehat{I}_{1,\alpha}(\epsilon)$	$\widehat{I}_{2,\alpha}(\epsilon)$
$[0.5318, 0.7990]$	$[0.3257, 0.5172]$	$[0.3536, 0.5934]$	$[0.3549, 0.5941]$	$[0.4159, 0.6635]$	$[0.3852, 0.6297]$

The above results indicates that the data $\widehat{Z}_{jn}(\epsilon)$ are not distributed according to the standard normal $N(0,1)$. We then attempt to verify if this situation is due to the influence of estimated first-step bandwidths $\widehat{b}_{0j}(\epsilon)$ on the variables $\widehat{Z}_{jn}(\epsilon)$. For this, we test the hypothesis $\widetilde{Z}_{jn}(\epsilon) \sim N(0,1)$ versus $\widetilde{Z}_{jn}(\epsilon) \not\sim N(0,1)$, where

$$\widetilde{Z}_{jn}(\epsilon) = \frac{\sqrt{n\widetilde{b}_j(\epsilon)}\left(\widetilde{f}_{jn}(\epsilon) - f(\epsilon)\right)}{\sqrt{f(\epsilon)\int K_1^2(v)dv}},$$

and $\widetilde{f}_{jn}(\epsilon)$ being at in the beginning of the subsection, in the comparison of the bias and variances. To perform the test, we consider T independent replications

$$\widetilde{Z^*}_{jk}(\epsilon) = \frac{\sqrt{n\widetilde{b}_j(\epsilon)}\left(\widetilde{f^*}_{jk}(\epsilon) - f(\epsilon)\right)}{\sqrt{f(\epsilon)\int K_1^2(v)dv}}, \quad k = 1, \ldots, T,$$

of the variables $\widetilde{Z}_{jn}(\epsilon)$. We denote by $\widetilde{D}_{jn}(\epsilon)$ the Kolmogorov-Smirnov statistic associated with the test. For our goodness of fit test, the null hypothesis is rejected with a level of significance α if $\sqrt{T}\widetilde{D}_{jn}(\epsilon) > K_\alpha$, where K_α satisfies

$$\mathbb{P}\left(\sqrt{T}\widetilde{D}_{jn}(\epsilon) \leq K_\alpha\right) = \mathbb{P}\left(\widetilde{D}_{jn}(\epsilon) \leq \frac{K_\alpha}{\sqrt{T}}\right) = 1 - \alpha.$$

The tables of critical values of the goodness of fit test to the standard normal variable can be found in the statistic literature. See, for example, Smirnov (1948), Miller (1956), Gibbons and Chakraborti (2003). Some of the results for the asymptotic approximations of the critical value K_α based on the ration $C_\alpha = K_\alpha/\sqrt{T}$ are :

$\mathbb{P}\left(\widetilde{D}_j(\epsilon) > C_\alpha\right)$	0.20	0.15	0.10	0.05	0.01
K_α	1.07	1.14	1.22	1.36	1.63

In Table 5.14, we give the values of the statistics $\sqrt{T}\widetilde{D}_{jn}(\epsilon)$ for $T = 100$, $n = 200$, $j = 1,2$ and $\epsilon = -1,0,1$. The results obtained here show that for the level $\alpha = 0.05$, the null hypothesis $\widetilde{Z}_{jn}(\epsilon) \sim N(0,1)$ is rejected, since $\sqrt{T}\widetilde{D}_{jn}(\epsilon) > K_\alpha = 1.36$, for all j and ϵ.

TABLE 5.14 – Values of the statistics $\sqrt{T}\widetilde{D}_{jn}(\epsilon)$ for $T = 100$, $j = 1,2$ and $\epsilon = -1,0,1$.

$\epsilon = -1$		$\epsilon = 0$		$\epsilon = 1$	
$\sqrt{T}\widetilde{D}_{1n}(\epsilon)$	$\sqrt{T}\widetilde{D}_{2n}(\epsilon)$	$\sqrt{T}\widetilde{D}_{1n}(\epsilon)$	$\sqrt{T}\widetilde{D}_{2n}(\epsilon)$	$\sqrt{T}\widetilde{D}_{1n}(\epsilon)$	$\sqrt{T}\widetilde{D}_{2n}(\epsilon)$
3.159609	3.354464	2.780215	2.676096	2.465744	1.890398

We now attempt to explain the non-validity of the hypothesis $\widetilde{Z}_{jn}(\epsilon) \sim N(0,1)$ by computing the empirical mean $\overline{\widetilde{\mu}}_{jn}(\epsilon)$ and the empirical variance $\overline{\widetilde{\sigma}}^2_{jn}(\epsilon)$ of the data $\widetilde{Z}_{jn}(\epsilon)$.

Table 5.15 – Values of the empirical means and variances of the $\widetilde{Z}_{jn}(\epsilon)$'s.

$\epsilon=-1$				$\epsilon=0$				$\epsilon=1$			
$\overline{\overline{\mu}}_{1n}(\epsilon)$	$\overline{\overline{\sigma}}^2_{1n}(\epsilon)$	$\overline{\overline{\mu}}_{2n}(\epsilon)$	$\overline{\overline{\sigma}}^2_{2n}(\epsilon)$	$\overline{\overline{\mu}}_{1n}(\epsilon)$	$\overline{\overline{\sigma}}^2_{1n}(\epsilon)$	$\overline{\overline{\mu}}_{2n}(\epsilon)$	$\overline{\overline{\sigma}}^2_{2n}(\epsilon)$	$\overline{\overline{\mu}}_{1n}(\epsilon)$	$\overline{\overline{\sigma}}^2_{1n}(\epsilon)$	$\overline{\overline{\mu}}_{2n}(\epsilon)$	$\overline{\overline{\sigma}}^2_{2n}(\epsilon)$
-0.4402	0.5322	-0.4044	0.4163	-0.4300	0.5743	-0.2880	0.4685	-0.3274	0.6124	-0.0773	0.5473

Table 5.15 shows that the estimated quantities $\overline{\overline{\mu}}_{jn}(\epsilon)$ and $\overline{\overline{\sigma}}^2_{jn}(\epsilon)$ are clearly different to 0 and 1. This should explain the rejection of the hypothesis $\widetilde{Z}_{jn}(\epsilon) \sim N(0,1)$, as seen above. Hence the results of our simulation study reveal that with the optimal step bandwitdhs $(\widehat{b}_{0j}(\epsilon), \widehat{b}_{1j}(\epsilon))$ and $\widetilde{b}_j(\epsilon)$, the variables $\widehat{Z}_{jn}(\epsilon)$ and $\widetilde{Z}_{jn}(\epsilon)$ are not distributed according to the standard distribution $N(0,1)$. However, the impact of the estimated optimal first-step bandwidths $\widehat{b}_{0j}(\epsilon)$ on the asymptotic normality of the variables $\widehat{Z}_{jn}(\epsilon)$ may not be so important as augured by the results obtained with the data $\widehat{Z}_{jn}(\epsilon)$.

5.4 Conclusion

The aim of this subsection was to analyze and compare the performances of the Kernel density estimator $\widehat{f}_{1n}$, based on the estimated residuals, and the ones of the integral Kernel estimator $\widehat{f}_{2n}$. Several aspects have been noticed in our simulation study. First, in the global framework, our numerical results show that the estimator $\widehat{f}_{1n}$ should be preferred to the estimator $\widehat{f}_{2n}$. The reason is that the optimal AISE of the latter estimator is much more higher than the one of the first estimator. For the evaluation of the bandwidths $(\widehat{b}_{1j}, \widehat{b}_{0j})$ that minimize the AISE of the estimators $\widehat{f}_{jn}$ $(j = 1, 2)$, our numerical results indicates that $\widehat{b}_{01}$ is much smaller than $\widehat{b}_{11}$, and that $\widehat{b}_{02}$ is approximately as small as $\widehat{b}_{12}$.

Next, for the pointwise study which is made at the points $\epsilon = -1, 0, 1$, we observe that $\widehat{f}_{1n}(\epsilon)$ dominates $\widehat{f}_{2n}(\epsilon)$ for $\epsilon = -1$ and $\epsilon = 1$ as well as for the ASE, the bias and the variance. Further, the ASE of the estimators $\widehat{f}_{jn}(\epsilon)$ are nearly the same as the ones of the estimators $\widetilde{f}_{jn}(\epsilon)$ for $\epsilon = 0$, and lower than the ASE of $\widetilde{f}_{jn}(\epsilon)$ for $\epsilon = -1$ and $\epsilon = 1$. In a semiparametric context, Müller, Schick and Wefelmeyer (2004) have shown that for the estimation of linear functionals of the error distribution, the estimators that use the estimated residuals may have a smaller asymptotic variance compared to the estimators based upon the true errors. Some of our simulation results suggest that a similar conclusion may hold when estimating the p.d.f. of regression residuals. In fact, for $\epsilon = 1$, the variances of the estimators $\widetilde{f}_{jn}(\epsilon)$ are higher than the ones of the estimators $\widehat{f}_{jn}(\epsilon)$. This shows that the estimation of the first-step bandwidth b_0 may have a positive influence when estimating $f(\epsilon)$.

The study of the asymptotic normality of the standardized variables $\widehat{Z}_{jn}(\epsilon)$ and $\widetilde{Z}_{jn}(\epsilon)$, based on the density estimators $\widehat{f}_{jn}(\epsilon)$ and $\widetilde{f}_{jn}(\epsilon)$, reveals that the data $\widehat{Z}_{jn}(\epsilon)$ and $\widetilde{Z}_{jn}(\epsilon)$ are normal, but are not distributed according to the standard normal variable. This means that the normal approximation of these variables by the normal $N(0, 1)$ is not satisfying for a small size of the samples ($n = 200$ in our framework). Therefore, it will be interesting, in a future works, to use the boostrap method for obtaining an alternative approximation of the considered variables. This will be one of the main aspects of the perspectives of our future researches, as illustrated at the end of this thesis.

Chapitre 6

Appendix

Abstract : This chapter contains some results which have an interest themselves and are used in Chapter 3 and Chapter 4. We begin with the Lyapounov Central Limit Theorem for triangular arrays which is used, for example, in the proof of Proposition 3.1 and Theorem 3.4. We also recall Theorem 1 and Theorem 2 in Einmahl and Mason (2005). These results are need in the validation of Lemma 3.1. We conclude by Theorem 2 in Whitlle (1960) and the Marcinkievicz-Zygmund inequality (see e.g Chow and Teicher 2003, p. 386) which are very useful for proving Lemma 3.10 and Lemma 4.12.

6.1 Lyapounov's Central Limit Theorem

For each integer $n \geq 1$, let $\{X_{1n}, X_{2n}, \ldots, X_{nn}\}$ be a collection of random variables such that $X_{1n}, X_{2n}, \ldots, X_{nn}$ are independent. Then $\{X_{1n}, X_{2n}, \ldots, X_{nn}\}$ is called a triangular array of independent variables.

Theorem 6.1. *(Lyapounov's Theorem)*
For all integer $n \geq 1$, assume that the variables X_{in}, $1 \leq i \leq n$, are independent with $\mathbb{E}[X_{in}] = 0$ for all i. Let $\alpha_n = \sqrt{\sum_{i=1}^{n} \mathrm{Var}(X_{in})}$. If there exists $\delta > 0$ such that

$$\lim_{n\to\infty} \alpha_n^{-(2+\alpha)} \sum_{i=1}^{n} \mathbb{E}\left[|X_{in}|^{2+\alpha}\right] = 0,$$

then

$$\frac{X_{1n} + X_{2n} + \ldots + X_{nn}}{\sqrt{\sum_{i=1}^{n} \mathrm{Var}(X_{in})}} \xrightarrow{d} N(0,1)$$

when $n \to \infty$.

This result can be found, for example, in Billingsley (1968, Theorem 7.3).

6.2 Uniform in bandwidth consistency of kernel-type function estimators

In this section, we give two results concerning the uniform in bandwidth consistency of kernel-type estimators, such that the density estimator and the regression function estimator. The results proposed here are established in Einmahl and Mason (2005). They are one of the keys of our main results in Chapter 3 and Chapter 4.

The first result we give concerns the Kernel density estimator. Let $X_1, X_2 \ldots, X_n$ be i.i.d $\mathbb{R}^d$, $d \geq 1$, valued random variables and assume that the common distribution function of the variables has a Lebesgue density function, which we denote by f. The Kernel density estimator of f based upon the sample $X_1, X_2, \ldots, X_n$, a Kernel function K and a bandwidth $0 < h = h(n) < 1$ is defined as

$$\widehat{f}_{n,h}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h^{1/d}}\right), \quad x \in \mathbb{R}^d.$$

For any function G defined and bounded on $\mathbb{R}^d$, we denote by $\|G\|_\infty$ the uniform norm of G such that

$$\|G\|_\infty = \sup_{x\in\mathbb{R}^d} |G(x)|\,.$$

The following theorem is proposed Einmahl and Mason (2005, p. 1382).

Theorem 6.2. *Assume that the Kernel function K is symmetric, continuous over $\mathbb{R}^d$ with support contained in $[-1/2, 1/2]^d$ and $\int K(x)dx = 1$. If the density function f is continuous and bounded on its support, then we have for any $C > 0$, with probability 1,*

$$\limsup_{n\to\infty} \sup_{C(\ln(n)/n)\leq h\leq 1} \|\widehat{f}_{n,h} - \mathbb{E}\widehat{f}_{n,h}\|_\infty = O\left(\frac{\sqrt{\ln(1/h) \vee \ln(\ln n)}}{nh}\right).$$

Remark : Choosing a sequence $h = h(n)$ satisfying $(nh/\ln n) \to \infty$ and $\ln(1/h)/\ln(\ln n) \to \infty$, one obtains, with probabilty 1,

$$\|\widehat{f}_{n,h} - \mathbb{E}\widehat{f}_{n,h}\|_\infty = O\left(\sqrt{(\ln(1/h))/(nh)}\right),$$

which is Theorem 1 of Giné and Guillou (2005).

The other kinds of kernel-type estimators treated by Einmahl and Mason is the regression Kernel estimators. For the illustration, consider i.i.d $(d+1)$-dimensional random vectors $(X, Y), (X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$, where the Y-variables are one-dimensional. We assume that X has a marginal Lebesgue density function f and that the regression function

$$m(x) = \mathbb{E}\left[Y \mid X = x\right], \quad x \in \mathbb{R}^d.$$

exists. Let $\widehat{m}_{n,h}(x)$ be the Nadaraya-Watson estimator of $m(x)$ with bandwidth $0 < h < 1$, that is,

$$\widehat{m}_{n,h}(x) = \frac{\sum_{i=1}^{n} Y_i K\left((x - X_i)/h^{1/d}\right)}{\sum_{i=1}^{n} K\left((x - X_i)/h^{1/d}\right)} .$$

With the above setup, we have the following uniform in bandwidth result. Let K and h be as in the previous section, and set

$$\overline{r}(x,h) = h^{-1}\mathbb{E}\left[YK\left(\frac{x-X}{h^{1/d}}\right)\right], \quad \overline{f}(x,h) = h^{-1}\mathbb{E}\left[K\left(\frac{x-X}{h^{1/d}}\right)\right].$$

For any subset I of $\mathbb{R}^d$, let I^ϵ denote its closed ϵ-neighborhood with respect to the maximum norm $|\cdot|_+$ on $\mathbb{R}^d$, that is, $|x|_+ = \max_{1\leq i\leq n} |x_i|$, $x \in \mathbb{R}^d$. Set further for any function $\psi : \mathbb{R}^d \to \mathbb{R}$, $\|\psi\|_I = \sup_{x\in\mathbb{R}} |\psi(x)|$.

Theorem 6.3. (Einmahl and Mason 2005, p. 1384)
Let I be a compact subset of $\mathbb{R}^d$ of $\mathbb{R}^d$ and assume that the Kernel function K satisfies the condition of Theorem 6.2. Suppose further that there exists an $\epsilon > 0$ so that f is continuous and strictly positive on $J := I^\epsilon$. If we assume that for some $p > 2$,

$$\sup_{z\in J} \mathbb{E}\left(|Y|^p \mid X = z\right) := \alpha < \infty,$$

we have for any $C > 0$ and $b_n \searrow 0$ with $\gamma = \gamma(p) = 1 - 2/p$,

$$\limsup_{n\to\infty} \sup_{C(\ln(n)/n)^\gamma \leq h\leq b_n} \|\widehat{m}_{n,h} - \overline{r}(\cdot,h)/\overline{f}(\cdot,h)\|_I = O\left(\frac{\sqrt{\ln(1/h) \vee \ln(\ln n)}}{nh}\right),$$

almost surely.

6.3 Bounds for the moments of linear forms in independent variables

The aim of this section is to propose absolute moments of linear forms in independent statistical variables. The first result we give here is established by Whitlle (1960, Theorem 2). Consider the linear form $L = \sum_{j=1}^{n} a_j\zeta_j$, where the ζ_j's are assumed to be independent mean-zero random variables, but not necessarily to be distributed identically. In what follows, we shall write

$$C(p) = \frac{2^{p/2}}{\sqrt{\pi}} \int_{-\infty}^{+\infty} |x|^p e^{-x^2} dx,$$

and $\gamma_j(p) = \left(\mathbb{E}\,|\zeta_j|^p\right)^{1/p}$, $p > 0$, provided that these quantities exist.

Theorem 6.4. (Whittle, 1960)
Then the following inequality is valid

$$\mathbb{E}\left(|L|^p\right) \leq 2^p C(p) \left(\sum_{j=1}^{p} \gamma_j^2(p)\right)^{p/2},$$

provided that $p \geq 2$ and the right-hand member exists. Moreover, if all the ζ_j have symmetric distributions, then the right-hand member may be divided by 2^p.

The second result we give is the Marcinkiewicz-Zygmund Inequality (See Chow and Teicher 2003, p.386). For any $p \geq 1$, let $\|\cdot\|_p$ denotes the L^p-norm, that is, $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$ for any random variable X such that $\mathbb{E}(|X|^p) < \infty$.

Theorem 6.5. Marcinkiewicz-Zygmund Inequality
If $\{X_n, n \geq 1\}$ are independent random variables with $\mathbb{E}[X_n] = 0$ for all n, then for every $p \geq 1$, there exist positive constant A_p and B_p depending only upon p for which

$$A_p \left\| \left(\sum_{j=1} X_j^2 \right)^{1/2} \right\|_p \leq \left\| \sum_{j=1} X_j \right\|_p \leq B_p \left\| \left(\sum_{j=1} X_j^2 \right)^{1/2} \right\|_p .$$

The proof of this Theorem can be found, for example, in Chow and Teicher (2003, p. 386).

Perspectives

Abstract

In this section, we sketch some perspectives for possible future researches. First, we have seen in our simulation study that the estimator of $f(\epsilon)$ introduced in Chapter 3 would be preferred to the one proposed in Chapter 4. However, it would be very interesting to compare the theoretical bias of the two estimators for determining the estimator that have to be used in a given context.

Our numerical results also reveal a curious situation : the estimator $\widehat{f}_{jn}(\epsilon)$ $(j = 1, 2)$ is sometimes more efficient than the estimator $\widetilde{f}_{jn}(\epsilon)$ when we are interested in their pointwise study. This situation comes from the evaluation of the second order of $\widehat{f}_{jn}(\epsilon)$, that is $\widehat{f}_{jn}(\epsilon) - \widetilde{f}_{jn}(\epsilon)$, which possibly allows to improve the performances of $\widehat{f}_{jn}(\epsilon)$. This curious siuation makes one to think that the term $\widehat{f}_{jn}(\epsilon) - \widetilde{f}_{jn}(\epsilon)$ is worth thinking about and deserved further consideration. We shall also attempt to obtain the uniform weak consistency for the difference $\widehat{f}_{jn}(\epsilon) - \mathbb{E}_n \widehat{f}_{jn}(\epsilon)$.

All the results proposed in this thesis are obtained in the case of a homoscedastic model. Then another axis for future researches will concern the extension of our results in a heteroscedastic framework, when the variance function depends upon the explanatory variable.

Résumé

Dans cette partie, nous donnons une esquisse des perspectives de recherche pour nos futurs travaux. D'abord, les résultats de nos simulations numériques montrent que l'estimateur de $f(\epsilon)$ introduit au Chapitre 3 devrait être preferé à celui défini au Chapitre 4. Cependant, il serait intéressant de comparer de façon théorique les biais des deux estimateurs. Ce sera l'un des problèmes sur lesquels nous nous pencherons dans nos recherches ultérieures.

Les résultats de nos simulations montrent également un point assez curieux : l'estimateur $\widehat{f}_{jn}(\epsilon)$ $(j = 1, 2)$ est parfois plus efficace que l'estimateur $\widetilde{f}_{jn}(\epsilon)$ lorsqu'on les étudie ponctuellement. Cette situation est due au second ordre de $\widehat{f}_{jn}(\epsilon)$, c'est à dire $\widehat{f}_{jn}(\epsilon) - \widetilde{f}_{jn}(\epsilon)$, qui permet éventuellement d'améliorer les performances de $\widehat{f}_{jn}(\epsilon)$. Ce deuxième ordre mériterait d'être étudié de façon plus poussée. Nous tenterons aussi d'obtenir des résultats de consistance uniforme pour

la différence entre $\widehat{f}_{jn}(\epsilon)$ et $\mathbb{E}_n\widehat{f}_{jn}(\epsilon)$.

Tous les résultats proposés dans cette thèse ont été obtenus dans un modèle de régression homoscédastique. Un autre axe de recherche pour nos futurs travaux sera de voir si des résultats comparables peuvent être obtenus dans le cas du modèle hétéroscédastique, où l'erreur du modèle dépend de la variable explicative.

Bibliographie

[1] Ahmad, I. & Li, Q. (1997). Testing symmetry of an unknown density function by kernel method. *Nonparam. Statistics.* **7**, 279–293.

[2] Altman, N. and Léger, C. (1995). Bandwidth selection for kernel distribution function estimation. *J. Statist. Plann. Inf.* **46**, n°2, 195–214.

[3] Akritas, M. G. & Van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scandinavian Journal of Statistics.* **28**, 549–567.

[4] Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika.* **68**, n°1, 326-328.

[5] Bai, J. (1994). Weak convergence of the sequential empirical processes in ARMA models. *Ann. Statist.* **22**, 2051–2061.

[6] Bashtannyk, D., M. and Hyndman, R., J. (2001). Bandwidth selection for conditional density estimation. *Comput. Statist. Data Anal.* **36**, n°3, 279–298.

[7] Bertrand-Retali, M. (1978). Convergence uniforme d'un estimateur de la densité par la méthode du noyau. *Rev. Roumaine Math. Pures et Appliquées.* **23**, n°3, 361-385.

[8] G. Biau, B. Cadre, B. Pelletier, D. Mason. (2009). Asymptotic normality in density support estimation. *Electron. J. Probab.*, 2617-2635.

[9] Biau, G., Cadre, B. and Pelletier, B. (2008). Exact rates in density support estimation. *J. Multivariate Anal.* **99**, 2185–2207.

[10] Bickel, P.J., Klaasssen, C.A.J., Ritov, Y. and Wellner, J.A. (1998). *Efficient and adaptive estimation for semi-parametric models.* Springer, New York.

[11] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley.

[12] Boldin, M. V. (1982). Estimation of the distribution of noise in an autoregressive scheme. *Theory Probab. Appl.* **27**, 866–871.

[13] Boldin, M. V. (1998). On residual empirical distribution functions in ARCH models with applications to testing and estimation. *Mitt. Math. Sem. Giessen* **235**, 49–66.

[14] Bosq, D. and Blanke, D. (2007). *Inference and prediction in large dimensions.* Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester Economica.

[15] Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction.* Lecture Notes in Staistics, 110, 2nd ed. Springer-Verlag, New York.

[16] Bosq, D., Lecoutre, J.P. (1987). *Théorie de l'estimation fonctionnelle.* Economica.

[17] Borovkov, A. (1987). *Statistique mathématique.* MIR, Moscou.

[18] Bowman, A., Hall, P. and Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika.* **85**, n°4, 799-808.

[19] Cadre, B. (2006). Kernel estimation of density levels sets. *J. Multiv. Anal.* **97**, n°4, 999–1023.

[20] Breimen, L., Meisel, W. and Purcell, E. (1977). Variable kernel estimate of multivariate densities. *Technometrics.* **19**, 135–144.

[21] Cheng, F. (2005). Asymptotic distributions of error density and distribution function estimators in nonparametric regression. *Journal of Statistical Planning and Inference* **128**, 327–349.

[22] Cheng, F., Sun, S. (2008). A goodness-of-fit test of the errors in nonlinear autoregressive time series models. *Statistics and Probability Letters* **78**, 50–59.

[23] Chow, Y., S. & Teicher, H. (2003). *Probability Theory : Independence, Interchangeability, Martingales.* Springer, 3rd ed.

[24] Conover, W., J. (1999). *Pratical Nonparametric Statistics.*, Third Edition, John Wliey & Sons, New York.

[25] Dallal, G., E. and Wilkinson, L. (1986). An analytic approximation to the distribution of Lilliefors' test statistic for normality. *The American Statistician.* **40**, 294–296.

[26] Collomb, G. (1979). Conditions nécessaires et suffisantes de convergence uniforme d'un estimator de régression, estimation des dérivées de la régression. *C.R. Acad. Sci. Paris Serie A.* **288**, 161–163.

[27] D'Agostino, R. and Stephens, M. (1986). *Goodness of fit techniques.*, M. Dekker, Inc., New York.

[28] De Gooijer, J., G. and Zerom, D. (2003). On conditional density estimation. *Statist. Neerlandica.* **57**, n°2, 159–176.

[29] Denby, L., Martin, D. (1979). Robust estimation of the first order autoregressive parameter. *J. Amer. Statist. Assoc.* **74**, 140–146.

[30] Dette, H. and von Lieres und Wilkau, C. (2001). Testing additivity by kernel based methods– What is a reasonable test ? *Bernoulli* **7**, 669–697.

[31] Dette, H., Kusi-Appiah, S., Neumeyer, N., (2002). Testing symmetry in nonparametric regression models. *Nonparam. Statistics.* **14(5)**, 477–494.

[32] Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Annals of Statistics.* **1)**, 279–290.

[33] Efromovich, S. (2005). Estimation of the density of the regression errors. *Annals of Statistics.* **33**, n°5, 2194–2227.

[34] Efromovich, S. (2007). Adaptive estimation of error density in nonparametric regression with small sample size. *Journal of Statistical Planning and Inference.* **137**, 363–378.

[35] Einmahl, U. & Mason, D. M. (2005). Uniform in bandwidth consistency of Kernel -type functions estimators. *Annals of Statistics.* **33**, 1380–1403.

[36] Escanciano, J., C. and Jacho-Chavez, D. (2010). $\sqrt{n}$-uniformly consistent density estimation in nonparametric regression models (submitted).
Einmahl, U. & Mason, D. M. (2005). Uniform in bandwidth consistency of Kernel -type functions estimators. *Annals of Statistics.* **33**, 1380–1403.

[37] Fan, J. and Yim, T., H. (2004). A crossvalidation method for estimating conditional densities. *Biometrika.* **91**, n° 4, 819–834.

[38] Feller, W. (1948). On the Kolmogorov-Smirnov Limit Theorems for Empirical Distributions. *Annals of Mathematical Statistics.* **19**, 177–189.

[39] Ferignac, P. (1962). Test de Kolmogorov-Smirnov sur la validité d'une fonction de distribution. *Revue de Statistique Appliquée.* **tome 10**, n° 4, 13–32.

[41] Fu, K., Yang, X. (2008). Asymptotics of Kernel error density estimators in nonlinear autoregressive models. *J. Math. Chem.* **44**, 831–838.

[41] Francq, C. and Tran, L., T. (2002). Nonparametric estimation of density, regression and dependence coefficients. *Journal of Nonparametric Statistics.* **14**, n°6, 729–747.

[42] Gibbons, J., D. and Chakraborti, S. (2003). *Nonparametric Statistical Inference.*, Fourth Edition, Revised and Expanded, Marcel Dekker, Inc.

[43] Giné, E. and Guillou A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. Henry Poincaré Prob. Statistics.* **38**, 907–921.

[44] Gozalo, P., L. and Linton, O., B. (2001). A nonparametric test of additivity in generalized nonparametric regression with estimated parameters. *J. Econometrics.* **104**, 1–48.

[45] Györfi, L. and Kohler, M. (2007). Nonparametric estimation of conditional distributions. *IEEE Trans. Inform. Theory.* **53**, n°5, 1872–1879.

[46] Hall, P., Racine, J. and Li, Q. (2004). Cross-validation and the estimation of conditional probability densities. *J. Amer. Statist. Assoc.* **99**, n° 468, 1015–1026.

[47] Hall, P., Carroll, R.J (1989). Variance function estimation in regression : The effect of estimating mean. *J. Roy. Statist. Soc. Ser. B,* **51**, 3–14.

[48] Härdle, W., Müller, M., Sperlich, S. and Werwatz, A. (2004). *Nonparametric and semiparametric models.* Springer Series in Statistics. Springer-Verlag, New-York.

[49] Hjort, N., L. and Jones, M., C. (1996). Locally parametric nonparametric density estimation. *Ann. Statist.* **24**, n°4, 1619–1647.

[50] Hössjer, O. and Ruppert, D. (1995). Asymptotics for the transformation kernel density estimator. *Ann. Statist.* **23**, n°4, 1198–1222.

[51] Huber, P., H. (1981). Robust estimation of a location parameter. *Annals Math. Statist.* **35**, 73–101.

[52] Huber, P., H. (1981). *Robust Statistics.* Wiley, New York.

[53] Hyndman, R., J. and Yao, Q. (2002). Nonparametric estimation and symmetry tests for conditional density functions. *J. Nonpar. Statist.* **14**, n°3, 259–278.

[54] Hyndman, R., J., Bashtannyk, D.,M. and Grunwald, G., K. (1996). Estimating and visualizing conditional densities. *J. Comput. Graph. Statist.* **5**, n°4, 315–336.

[55] Jones, M.C., Marron, J.S. and Sheather., S.J. (1996). A brief survey of bandwidth selection for density estimation. *J. Amer. Statist. Assoc.* **91**, 401–407.

[56] Klimko, L.A., Nelson, P.I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.* **6**, 629–642.

[57] Kolmogorov, A. (1933) Grundbegriffe der Wahrscheinlichkeitsrechnung. Berlin : Julius Springer.
Translation : Kolmogorov, Andrey (1956) in Foundations of the Theory of Probability, 2nd ed., New York : Chelsea. pp. 50–51.

[58] Koul, H. L. (1992). *Weighted empiricals and linear models.* IMS, Hayward, CA.

[59] Koul, H. L., Sakhanenko, L. (2005). Goodness-of-fit testing in regression : A finite sample comparison of boostrap methodology and Khmaladze transformation. *Statistics and Probability Letters.* **74**, 290–302.

[60] Lehmann, E. L. and Romano, J., P. (1998). *Testing Statistical Hypotheses.*, Third Edition, Springer.

[61] Lejeune, M. and Sarda, P. (1992). Smooth estimators of distribution and density functions. *Computational Statistics and Data Analysis.* **14**, n° 4, 457–471.

[62] Lepski, O.V. and Spokoiny, V.G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Annals of Statistics* **25**, 2512–2546.

[63] Loftsgaarden, D.O. and Quesenberry, C.P. (1965). A nonparametric estimate of a multivariate density function. *Ann. Math. Statist.* **36**, 1049–1051.

[64] Loynes, R.M., (1980). The empirical distribution function of residuals from generalised regression. *Annals of Statistics* **8**, 285–298.

[65] Mammen, M. (1996). Empirical process of residuals for high-dimensional linear models. *Ann. Statist.* **24**, 307–335.

[66] Massey, F., J., Jr (1951). The Komogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association.* **46**, 253 (March 1951), 68–78.

[67] Miller, L., H. (1956). Table of percentage points of Kolmogorov Statistics. *Journal of the American Statistical Association.* **51**, 273 (March 1956), 111–121.

[68] Müller, U. U., Schick, A. et Wefelmeyer, W. (2004). Estimating linear functionals of the error distribution in nonparametric regression. *J. Statist. Plann. Inference.* **119**, 75–93.

[69] Müller, H.G. and Stadmüller, U. (1987). Variable bandwidth kernel estimators of regression curves. *Annals of Statistics* **15**, 182–201.

[70] Nadaraya, E., A. (1964). On a regression estimate. *Teor. Verojatnost. i Primenen.* **9**, 157–159.

[71] Nadaraya, E. A. (1989). *Nonparametric estimation of probability densities and regression curves.* Kluwer Academic Publishers.

[72] Neumeyer, N., Dette, H., Nagel, E-R. (2005). A note on testing symmetry of the error distribution in linear regression models. *Nonparametric Statistics* **17**, n°6, 697–715.

[73] Neumeyer, N. and Dette, H. (2007). Testing for symmetric error distribution in nonparametric regression models. *Statistica Sinica* **17**, 775–795.

[74] Neumeyer, N. and Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J. Multiv. Analysis.* **101**, 1067–1078.

[75] Parzen, E. (1962). On estimation of a probability distribution function and mode. *Ann. Math. Statist.* **33**, 1065–1076.

[76] Pinsker, M.S. (1980). Optimal filtering of a square integrable signal in Gaussian white noise. *Problems Inform. Transmission* **16**, 52–68.

[77] Plancade, S. (2008). Nonparmetric estimation of the density of the regression noise. *C.R. Acad. Sci. Paris* **Ser. I 346**, 461–466.

[78] Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

[79] Portnoy, S. (1986). Asymptotic behavior of the empirical distribution of M-estimated residuals from a regression model with many parameters. *Ann. Statist.,* **14**, 1152–1170.

[80] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of density function. *Ann. Math. Statist.,* **27**, 832–837.

[81] Roussas, G. (1967). Nonparametric Estimation in Markov processes. *Technical Repport* **110**, *Univ. of Wisconsin, Madison.*

[82] Roussas, G. (1991). Estimation of transition distribution function and its quantiles in Markov processes : strong consistency and asymptotic normality. In : *Nonparametric Functional Estimation and Related Topics*, pp. 463-474. *Kluwer, Dordrecht.*

[83] Sarda, P., Vieu, P. and Youndjé, E. (1996). Choix de paramètres de lissage en estimation de densité conditionnelle. *Publi. de l'Inst. de Statist. de Paris,* **XXXVIII**, 57–80.

[84] Sarda, P. (1993). Smoothing parameter selection for smooth distribution functions. *J. Statist. Plann. Inference,* **35**, n°1, 65–75.

[85] Scott, W., S. (1992). *Multivariate density estimation.* Wiley.

[86] Serfling, R.J. (2002). *Approximation Theorems of Mathematical Statistics.* Paperback Edition, Wiley.

[87] Shorack, G. R. (1984). Empirical and rank processes of observations and residuals. *Canad. J. Statist.,* **12**, 319–332.

[88] Shorack, G.,R. and Wellner, J., A. (1986).*Empirical Processes With Applications to Statistics*, John Wiley & Sons Inc.

[89] Smirnov, N. (1948). Table for estimating goodness of fit of empirical distributions. *Annals of Mathematical Statistics,* **19**, 279–281.

[90] Staniswalis, J. G. (1989). Local bandwidth selection for kernel estimates. *Journal of American Statistical Association,* **84**, 284–288.

[91] Stute, W. (1982). The law of the iterated logarithm for kernel density estimators. *Annals of Probability,* **10**, 414–422.

[92] Stone, C. J. (1977). Consistent nonparametric regression. *Annals of Statistics,* **5**, 595–645.

[93] Tassi, P. (1985). *Méthodes Statistiques.*, Ed. Economica.

[94] Terrell, G.R. and Scott, D.W. (1992). Variable kernel density estimation. *Annals of Statistics,* **20**, n°3, 1236–1265.

[95] Thode Jr, H., C. (2002). *Testing for normality.*, Marcel Dekker, New York.

[96] Van Der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge Univ. Press.

[97] Van Der Vaart, A. W., Wellner, J. A. (1996). *Weak Convergence and Empirical Processes : With Applications to Statistics.* Springer, New York.

[98] Van Keilegom, I. and Veraverbeke, N. (2002). Density and hazard estimation in censored regression models. *Bernoulli,* **8(5)**, 607–625.
Watson, G., S. (1964). Smooth regression analysis. *Sankhyā*, **Ser. A 26**, 359–372.

[99] Wand, M., P., Jones, M., C. (1995). *Kernel Smoothing.* Chapman & Hall/CRC.

[100] Wang, L., Brown, L. D., Cai, T. T., Levine, M. (2008). Effect of mean on variance function estimation in nonparametric regression. *Ann. Statist.,* **36**, 646–664.

[101] Watson, G., S. (1964). Smooth regression analysis. *Sankhyā*, **Ser. A 26**, 359–372.

[102] Whittle, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. *Theory of Probability and its Applications,* **5**, 302–305.

[103] Youndjé, E. (1996). Propriétés de convergence de l'estimateur à noyau de la densité conditionnelle. *Revue Roumaine de Mathématiques Pures et Appliquées,* **41**, 535–566.

[104] Youndjé, E., Sarda, P. and Vieu, P. (1994). Validation croisée pour l'estimation nonparamétrique de la densité conditionnelle. *Publ. Inst. Statist. Univ. Paris,* **38**, n°1, 57–80.